

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 11, 2001, 12:58:02 ; Search time 8.64 Seconds

(without alignments)  
41.115 Million cell updates/sec

Title: US-09-722-440-8

Perfect score: 70

Sequence: 1 CAMELEMPCA 11

Scoring table: BLOSUM62

Gapop 10.0 , Capext 0.5

arched: 88757 seqs, 32294092 residues

Total number of hits satisfying chosen parameters: 88757

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : SwissProt\_39:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	41	58.6	546	1 BGAM_HUMAN	P16279 homo sapien
2	41	58.6	677	1 BGAL_HUMAN	P16279 homo sapien
3	40	57.1	374	1 COSA_YEAST	P52924 saccharomyc
4	39	55.7	419	1 VSYL_TRYCO	P20948 trypanosoma
5	38	54.3	310	1 KITH_HSVTF	P13157 turkey herp
6	38	54.3	350	1 KITH_HSVTF	P25987 turkey herp
7	38	54.3	630	1 PRLR_ORENI	O91513 oreochromis
8	38	54.3	1101	1 GUNC_CELFI	P14090 cellulomona
9	37	52.9	66	1 UR90_HCVYA	P16796 human cytom
10	37	52.9	229	1 ABME_MOUSE	P51908 mus musculu
11	37	52.9	229	1 ABME_MOUSE	P38483 rattus norv
12	37	52.9	236	1 ABME_HUMAN	P41238 homo sapien
13	37	52.9	236	1 ABME_HUMAN	P41238 homo sapien
14	37	52.9	574	1 CO9_ONCMY	P06682 oncorhynch
15	37	52.9	614	1 NTBE_MOUSE	P31651 mus musculu
16	37	52.9	614	1 NTBE_MOUSE	P48056 rattus norv
17	37	52.9	614	1 NTBE_MOUSE	P38811 saccharomyc
18	36.5	52.1	1039	1 ITAB_HUMAN	P08514 homo sapien
19	36	51.4	176	1 CRIZ_ERMHE	O01332 erwina her
20	36	51.4	263	1 RYPL_SALTY	O92466 salmonele
21	36	51.4	263	1 RYPL_SALTY	O92466 salmonele
22	36	51.4	316	1 MUCB_PSEAE	P38108 pseudomonas
23	36	51.4	366	1 YHKO_YEAST	P10259 mouse mamma
24	36	51.4	366	1 YHKO_YEAST	P10259 mouse mamma
25	36	51.4	366	1 YHKO_YEAST	P10259 mouse mamma
26	36	51.4	366	1 YHKO_YEAST	P10259 mouse mamma
27	35	50.0	124	1 YBA4_YEAST	P03374 mouse mamma
28	35	50.0	182	1 TERS_BPP21	P31941 homo sapien
29	35	50.0	246	1 MOC_MOUSE	P36694 bacterioph
30	35	50.0	301	1 OPSD_CAMHU	O61885 mus musculu
31	35	50.0	301	1 OPSD_CAMHU	O61885 mus musculu
32	35	50.0	420	1 YMS7_YEAST	O18312 cambarus hu
33	35	50.0	511	1 GUNB_PSEFL	O03694 saccharomyc
					P18126 pseudomonas

34	35	50.0	590	1 GUAA_MYCLE	P46810 mycobacteri
35	35	50.0	2491	1 MPRI_HUMAN	P11717 homo sapien
36	34.5	49.3	534	1 YN22_CAHEL	O21339 caenorhabdi
37	34.5	49.3	804	1 YBAP_ECOLI	P77504 escherichia
38	34	48.6	67	1 ATP8_RABIT	O79431 oryctolagus
39	34	48.6	88	1 TOXK_WILMR	P12501 simlan immu
40	34	48.6	100	1 POL_SIVAS	P12501 simlan immu
41	34	48.6	105	1 NIPM_BOVIN	O02379 bos taurus
42	34	48.6	105	1 NIPM_BOVIN	O02379 bos taurus
43	34	48.6	228	1 RNS4_PYRPT	O40966 pyrus pyrif
44	34	48.6	241	1 YFBB_ECOLI	P37355 escherichia
45	34	48.6	245	1 CCMC_ECOLI	P33929 escherichia

## ALIGNMENTS

```

RESULT 1
ID BGAM_HUMAN STANDARD: PRT; 546 AA.
AC P16279;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE BETA-GALACTOSIDASE-RELATED PROTEIN PRECURSOR.
GN GUBI.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
RN [1]
RP SEQUENCE FROM N.A., AND PLACENTAL, PARTIAL SEQUENCE.
RC TISSUE-TESTIS;
RX MEDLINE-90062209; Pubmed-2511208;
RA Morreau H., Galjart N.J., Gillemaus N., Willemsen R.,
RA van der Horst G.T.J., D'Azco A.;
RT "Alternative splicing of beta-galactosidase mRNA generates the
RT classic lysosomal enzyme and a beta-galactosidase-related protein.";
RL J. Biol. Chem. 264:20655-20663(1989).
CC -1- FUNCTION: THIS PROTEIN HAS NO CATALYTIC ACTIVITY.
CC -1- ALTERNATIVE PRODUCTS: BETA-GALACTOSIDASE AND BETA-
CC RELATED PROTEIN ARE PRODUCED BY ALTERNATIVE SPLICING OF THE SAME
CC GENE.
CC -1- SIMILARITY: BELONGS TO FAMILY 35 OF GLYCOSYL HYDROLASES.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M27508; AAA35599.1; -.
CC PIR; B32688; B32688.
CC INTERPRO: IPR001944; -.
CC PFMAM; PF01301; Glyco_hydro_35; 1.
CC
CC SIGNAL: Alternative splicing; Glycoprotein.
CC
CC SIGNAL 1 22
CC CHAIN 24 546
CC CARBOHYD 26 26
CC CARBOHYD 116 116
CC CARBOHYD 333 333
CC CARBOHYD 367 367
CC CARBOHYD 411 411
CC CARBOHYD 414 414
CC CARBOHYD 424 424
CC
CC SEQUENCE 546 AA; 60551 MW; 2B1A3EDAF95966C CRC64;

```

Query Match 58.6%; Score 41; DB 1; Length 546;  
Best Local Similarity 75.0%; Pred. No. 12;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 4 ELEMPCA 11  
 RL 1111 11:  
 DB 489 ELEMAPCS 496

RESULT 2  
 BGAL\_HUMAN STANDARD; PRT; 677 AA.  
 AC P16278;  
 DT 01-AUG-1990 (Rel. 15, Created)  
 DT 01-AUG-1990 (Rel. 15, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE BETA-GALACTOSIDASE PRECURSOR (EC 3.2.1.23) (LACTASE) (ACID BETA-GALACTOSIDASE).  
 GN GBL1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 RN [1]  
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 RC TISSUE=TESTIS;  
 RX MEDLINE=90062209; Pubmed=2511208;  
 RA Morreau H., Galjart N.J., Gillemans N., Willemsen R., van der Horst G.T.J., D'Azzo A.,  
 RT "Alternative splicing of beta-galactosidase mRNA generates the classic lysosomal enzyme and a beta-galactosidase-related protein.",  
 RL J. Biol. Chem. 264:20655-20663(1989).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=90262647; Pubmed=2111707;  
 RA Yamamoto Y., Hake C.A., Martin B.M., Kretz K.A., Ahern-Rindel A.J., Naylor S.L., Mudd M., O'Brien J.S.,  
 RT "Isolation, characterization, and mapping of a human acid beta-galactosidase cDNA.",  
 RL DNA Cell Biol. 9:119-127(1990).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=PLACENTA;  
 RX MEDLINE=89061717; Pubmed=3143362;  
 RA Oshima A., Tsuji A., Nagao Y., Sakuraba H., Suzuki Y.,  
 RT "Cloning, sequencing, and expression of cDNA for human beta-galactosidase.",  
 RL Biochem. Biophys. Res. Commun. 157:238-244(1988).  
 RN [4]  
 RP VARIANTS MOROITO B LEU-273; HIS-482 AND CYS-509.  
 RX MEDLINE=92026088; Pubmed=1928092;  
 RA Oshima A., Yoshida K., Shimoto M., Fukuhara Y., Sakuraba H., Suzuki Y.,  
 RT "Human beta-galactosidase gene mutations in morquio B disease.",  
 RL Am. J. Hum. Genet. 49:1091-1093(1991).  
 RN [5]  
 RP VARIANTS GMI CYS-49; THR-51 AND CYS-201.  
 RX MEDLINE=91353572; Pubmed=1909089;  
 RA Nishimoto J., Namba E., Inui K., Okada S., Suzuki K.,  
 RT "GMI-gangliosidosis (genetic beta-galactosidase deficiency): identification of four mutations in different clinical phenotypes among Japanese patients.",  
 RL Am. J. Hum. Genet. 49:566-574(1991).  
 RN [6]  
 RP VARIANTS GMI THR-51; ARG-123; CYS-201; CYS-316 AND GLN-457.  
 RX MEDLINE=91328151; Pubmed=1907800;  
 RA Yoshida K., Oshima A., Shimoto M., Fukuhara Y., Sakuraba H., Yanagisawa N., Suzuki Y.,  
 RT "Human beta-galactosidase gene mutations in GMI-gangliosidosis: a common mutation among Japanese adult/chronic cases.",  
 RL Am. J. Hum. Genet. 49:435-442(1991).  
 RN [7]  
 RP VARIANT GMI HIS-482.  
 RX MEDLINE=93138608; Pubmed=1487238;  
 RA Mosna G., Fatore S., Tobiello G., Brocca S., Trubia M., Gianazza E., Gatti R., Danesino C., Minelli A., Plantanida M.,  
 RT "A homozygous missense arginine to histidine substitution at position 482 of the beta-galactosidase in an Italian infantile GMI-

RT gangliosidosis patient.",  
 RL Hum. Genet. 90:247-250(1992).  
 RN [8]  
 RP VARIANTS GMI CYS-208; ARG-578; HIS-590 AND GLY-632.  
 RX MEDLINE=94027054; Pubmed=8213816;  
 RA Boustany R.-M., Qian W.-H., Suzuki K.,  
 RT "Mutations in acid beta-galactosidase cause GMI-gangliosidosis in American patients.",  
 RL Am. J. Hum. Genet. 53:881-888(1993).  
 RN [9]  
 RP VARIANT GMI MET-82.  
 RX MEDLINE=94256487; Pubmed=8198123;  
 RA Chakraborty S., Rafi M.A., Wenger D.A.,  
 RT "Mutations in the lysosomal beta-galactosidase gene that cause the adult form of GMI gangliosidosis.",  
 RL Am. J. Hum. Genet. 54:1004-1013(1994).  
 RN [10]  
 RP VARIANTS MOROITO B HIS-83 AND CYS-482.  
 RX MEDLINE=96049832; Pubmed=7586649;  
 RA Ishii N., Oohira T., Oshima A., Sakuraba H., Endo F., Matsuda I., Sukegawa K., Orit T., Suzuki Y.,  
 RT "Clinical and molecular analysis of a Japanese boy with Morquio B disease.",  
 RL Clin. Genet. 48:103-108(1995).  
 RN [11]  
 RP VARIANTS GMI HIS-59; ASN-591 AND CYS-591.  
 RA Gatti R., Taddeucci G., Ricci R., D'Azzo A., Zammarchi E.,  
 RT "Identification of new mutations in six Italian patients with a variant form of infantile GMI-gangliosidosis with severe cardiomyopathy.",  
 RL Am. J. Hum. Genet. 61:A258-A258(1997).  
 RN [12]  
 RP VARIANTS GMI H-59;S-121;C-208;M-240;N-491,6 VARIANTS P-10;C-521;G-532.  
 RX MEDLINE=99268417; Pubmed=10338095;  
 RA Silva C.M.D., Severini M.H., Sopaesa A., Coelho J.C., Zaha A., d'Azzo A., Gugliani R.,  
 RT "Six novel beta-galactosidase gene mutations in Brazilian patients with GMI-gangliosidosis.",  
 RL Hum. Mutat. 13:401-409(1999).  
 CC -I- FUNCTION: CLEAVES BETA-LINKED TERMINAL GALACTOSYL RESIDUES FROM GALACTOSIDES, GLYCOPROTEINS, AND GLYCOSAMINOGLYCAN.  
 CC -I- CATALYTIC ACTIVITY: HYDROLYSIS OF TERMINAL, NON-REDUCING BETA-D-GALACTOSE RESIDUES IN BETA-D-GALACTOSIDES.  
 CC -I- SUBCELLULAR LOCATION: LYSOSOMAL.  
 CC -I- ALTERNATIVE PRODUCTS: BETA-GALACTOSIDASE AND BETA GALACTOSIDASE-RELATED PROTEIN ARE PRODUCED BY ALTERNATIVE SPLICING OF THE SAME GENE.  
 CC -I- DISEASE: DEFECTS IN GBL1 ARE THE CAUSE OF GMI-GANGLIOSIDOSIS, AN AUTOSOMAL RECESSIVE DISORDER WITH THREE MAJOR CLINICAL PHENOTYPES THAT ARE DISTINGUISHED ACCORDING TO THE AGE OF ONSET AND SEVERITY OF SYMPTOMS: INFANTILE, JUVENILE AND ADULT. THE INFANTILE FORM IS RAPIDLY PROGRESSIVE AND RESULTS IN SEVERE CENTRAL NERVOUS SYSTEM DEGENERATION AND VISCEROMEGALY, WITH DEATH USUALLY BETWEEN THE FIRST AND SECOND YEAR.  
 CC -I- DISEASE: DEFECTS IN GBL1 ARE THE CAUSE OF MUCOPOLYSACCHARIDOSIS IV B (OR MOROITO B SYNDROME) WHICH IS CHARACTERIZED BY SEVERE BONE DEFORMITIES WITHOUT CNS INVOLVEMENT.  
 CC -I- SIMILARITY: BELONGS TO FAMILY 35 OF GLYCOSYL HYDROLASES.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: M27507; AAA51819.1; -  
 DR EMBL: M34423; AAA51823.1; -  
 DR EMBL: M22590; AAA51822.1; -  
 DR PIR: A32611; A32611.  
 DR PIR: A31673; A31673.

```

DR PIR: A32688; A32688.
DR PIR: B37086; B37086.
DR MIM: 230500; -.
DR MIM: 230600; -.
DR MIM: 230650; -.
DR MIM: 253010; -.
DR INTERPRO: IPR001944; -.
DR PIR: PF01301; Glyco_hydro.35; 1.
DR PRINTS: PR00742; GLHYDRLASE35.
DR PROSITE: PS01182; GLYCOSYL_HYDROL_F35; 1.
KW Hydrolase; Glycosidase; Lysosome; Signal; Alternative splicing;
KW Glycoprotein; Polymorphism; Disease mutation.
FT SIGNAL 1 23
FT PROPEP 24 28
FT CHAIN 29 677
FT ACT_SITE 188 188
FT ACT_SITE 268 268
FT CARBOHYD 26 26
FT CARBOHYD 247 247
FT CARBOHYD 464 464
FT CARBOHYD 498 498
FT CARBOHYD 542 542
FT CARBOHYD 545 545
FT CARBOHYD 555 555
FT VARIANT 10 10
FT VARIANT 49 49
FT VARIANT 51 51
FT VARIANT 59 59
FT VARIANT 82 82
FT VARIANT 83 83
FT VARIANT 121 121
FT VARIANT 123 123
FT VARIANT 201 201
FT VARIANT 208 208
FT VARIANT 240 240
FT VARIANT 273 273
FT VARIANT 316 316
FT VARIANT 457 457
Query Match 58.6%; Score 41; DB 1; Length 677;
Best Local Similarity 75.0%; Pred. No. 15;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 4 ELEWLPKA 11
DB 620 ELEWAPCS 627
RESULT 3
COSM_YEAST

```

```

ID COSM_YEAST STANDARD; PRT; 374 AA.
AC P52924;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE COS10 PROTEIN.
GN COS10 OR YNR075W OR N3820.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC Saccharomycetaceae; Saccharomyces.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=5286C;
RX MEDLINE=97060021; Pubmed=8904342;
RT Levesque H., Lepingle A., Nicaud J.-M., Gaillardin C.;
RT *Sequencing of a 9.2 kb telomeric fragment from the right arm of
RT Saccharomyces cerevisiae chromosome XIV.*;
RT Yeast 12:289-295(1996).
RL [2]
RP SEQUENCE FROM N.A.
RA Andre B., Iraqi Housaini I., Urrestarazu L.A., Vissers S.;
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE DUP/COS FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-slb.ch/announce/
CC or send an email to license@slb.ch).
CC -----
CC EMBL: X86790; CAA60488.1; -.
CC EMBL: Z71690; CAA96359.1; -.
CC DR SGD: S0005358; COS10.
CC DR INTERPRO: IPR001142; -.
CC DR PFAM: PF00674; DUP: 2.
CC KW Transmembrane; Multigene family.
FT TRANSMEM 42 62
FT TRANSMEM 63 83
FT TRANSMEM 244 264
FT TRANSMEM 374 AA; 43866 MW; 7A7195A782065B5F CRC64;
SQ SEQUENCE
QY 3 WELEWLPKA 11
DB 45 WSLSWLPKA 53
Query Match 57.1%; Score 40; DB 1; Length 374;
Best Local Similarity 66.7%; Pred. No. 12;
Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 4 ELEWLPKA 11
DB 620 ELEWAPCS 627
RESULT 4
VSYL_TRYCO STANDARD; PRT; 419 AA.
ID VSYL_TRYCO
AC P20948;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE VARIANT SURFACE GLYCOPROTEIN YNAT 1.1 PRECURSOR (VSG).
OS Trypanosoma congolense.
OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=87185370; Pubmed=3567147;
RA Strickler J.E., Binder D.A., L'Italien J.J., Shimamoto G.T.,
RA Walt S.W., Dalheim L.J., Novotny J., Radding J.A., Konigsberg W.H.,
RA Armstrong M.Y.K., Richards F.F., Lalor T.M.;
RT "Trypanosoma congolense: structure and molecular organization of the
RT surface glycoproteins of two early bloodstream variants.";
RT Biochemistry 26:796-805(1987).

```

CC -1- FUNCTION: VSG FORMS A COAT ON THE SURFACE OF THE PARASITE. THE  
CC TYPANOSOME EVADES THE IMMUNE RESPONSE OF THE HOST BY EXPRESSING  
CC A SERIES OF ANTIGENICALLY DISTINCT VSGS FROM AN ESTIMATED 1000  
CC VSG GENES.  
CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR.  
CC A SOLUBLE FORM IS RELEASED FROM RUPTURED CELLS BY THE ACTION OF A  
CC PI-PLC.  
CC -1- SIMILARITY: 25% TO VARIANT SURFACE GLYCOPROTEIN YNAT 1.3.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL, M15112; AAA30298.1; -.  
CC PIR; A27539; A27539.  
CC Glycoprotein; Antigen; Trypanosomiasis; GPI-anchor; Membrane; Signal.  
CC SIGNAL 1 28  
CC CHAIN 29 400 VARIANT SURFACE GLYCOPROTEIN YNAT 1.1.  
CC PROPEP 401 419 HYDROPHOBIC, REMOVED DURING MATURATION.  
CC LIPID 400 400 GPI-ANCHOR.  
CC CARBOHYD 82 82 N-LINKED (GLCNAC. . .).  
CC CARBOHYD 358 358 N-LINKED (GLCNAC. . .). (HIGH MANNOSE).  
CC CARBOHYD 379 379 N-LINKED (GLCNAC. . .). (HIGH MANNOSE).  
CC SEQUENCE 419 AA; 44767 MW; F1BC1F8C639CDF02 CRC64;

Query Match 55.7%; Score 39; DB 1; Length 419;  
Best Local Similarity 75.0%; Pred. No. 20;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 4 ELEWPCA 11  
1:|||||  
DB 55 EEWLPCA 62

RESULT 5  
KITH\_HSVTF STANDARD; PRT; 310 AA.  
AC P1315;  
DT 01-JAN-1990 (Rel. 13, Created)  
DT 01-JAN-1990 (Rel. 13, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE THYMIDINE KINASE (EC 2.7.1.21).  
GN TK.  
OS Turkey herpesvirus (strain FC126).  
US Viruses; dsDNA viruses, no RNA stage; Herpesviridae.  
[1]  
SEQUENCE FROM N.A.  
RX MEDLINE=69259069; PubMed=2724415;  
RA Martin S.L., Aparicio D.I., Bandyopadhyay P.K.;  
RT Genetic and biochemical characterization of the thymidine kinase  
RT gene from herpesvirus of turkeys.";  
RL J. Virol. 63:2847-2852(1989).  
CC -1- CATALYTIC ACTIVITY: ATP + THYMIDINE = ADP + THYMIDINE  
CC 5'-PHOSPHATE.  
CC  
CC -1- SIMILARITY: BELONGS TO THE HERPESVIRUSES THYMIDINE KINASE FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL, M26659; AAA46109.1; -.  
CC PIR; A33346; KIBETH.  
CC HSSP; P03176; IKIM.  
CC INTERPRO: IPR001889; -.

DR PFAM: PF00693; TK\_herpes; 1.  
KW Transferase; Kinase; DNA synthesis; ATP-binding.  
FT NP\_BIND 17 24 ATP (PROBABLE).  
SQ SEQUENCE 310 AA; 35512 MW; 9E796BD643EB6F38 CRC64;

Query Match 54.3%; Score 38; DB 1; Length 310;  
Best Local Similarity 83.3%; Pred. No. 21;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 WELEWL 8  
1:|||||  
DB 220 WELEWL 225

RESULT 6  
KITH\_HSVTU STANDARD; PRT; 350 AA.  
AC P25987;  
DT 01-MAY-1992 (Rel. 22, Created)  
DT 01-MAY-1992 (Rel. 22, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE THYMIDINE KINASE (EC 2.7.1.21).  
GN TK.  
OS Turkey herpesvirus.  
US Viruses; dsDNA viruses, no RNA stage; Herpesviridae.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=90063552; PubMed=2555435;  
RA Scott S.D., Ross N.L.J., Binns M.M.;  
RT "Nucleotide and predicted amino acid sequences of the Marek's disease  
RT virus and turkey herpesvirus thymidine kinase genes; comparison with  
RT thymidine kinase genes of other herpesviruses.";  
RL J. Gen. Virol. 70:3055-3065(1989).  
CC -1- CATALYTIC ACTIVITY: ATP + THYMIDINE = ADP + THYMIDINE  
CC 5'-PHOSPHATE.  
CC  
CC -1- SIMILARITY: BELONGS TO THE HERPESVIRUSES THYMIDINE KINASE FAMILY.  
CC  
CC PIR; A33375; KIBEC.  
DR HSSP; P03176; IKIM.  
DR INTERPRO: IPR001889; -.  
DR PFAM: PF00693; TK\_herpes; 1.  
KW Transferase; Kinase; DNA synthesis; ATP-binding.  
FT NP\_BIND 17 24 ATP (PROBABLE).  
SQ SEQUENCE 350 AA; 39968 MW; AD3B8F96E431EB83 CRC64;

Query Match 54.3%; Score 38; DB 1; Length 350;  
Best Local Similarity 83.3%; Pred. No. 24;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 WELEWL 8  
1:|||||  
DB 220 WELEWL 225

RESULT 7  
PRLR\_ORENI STANDARD; PRT; 630 AA.  
AC Q91513;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 15-JUL-1998 (Rel. 36, Last annotation update)  
DE PROLACTIN RECEPTOR PRECURSOR (PRL-R).  
GN PRLR.  
OS Oreochromis niloticus (Nile tilapia) (Tilapia nilotica).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
OC Acanthomorpha; Acanthopterygii; Perciformes; Labroidae;  
OC Cichlidae; Oreochromis.  
RN [1]  
RP SEQUENCE FROM N.A.  
RP TISSUE=KIDNEY;  
RC MEDLINE=95320210; PubMed=7597076;  
RX

RA Sandra O., Sohm F., de Luze A., Prunet P., Edey M., Kelly P.A.;  
 RT "Expression cloning of a cDNA encoding a fish prolactin receptor."  
 RL Proc. Natl. Acad. Sci. U.S.A. 92:6037-6041(1995).  
 CC  
 CC -1- FUNCTION: THIS IS A RECEPTOR FOR THE ANTERIOR PITUITARY HORMONE  
 CC PROLACTIN.  
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.  
 CC -1- SIMILARITY: BELONGS TO THE CYTOKINE FAMILY III-LIKE DOMAINS.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: L34783; AAA98997.1; -  
 CC HSSP: P16471; 1BR3.  
 CC INTERPRO: IPR000950; -  
 CC INTERPRO: IPR001777; -  
 CC INTERPRO: IPR002465; -  
 CC PFAM: PF00041; fn3; 2.  
 CC DR PROSITE: PS01352; HEMATOPO. REC. L.F1; 1.  
 CC KW Receptor; Transmembrane; Glycoprotein; Signal; Repeat.  
 CC FT SIGNAL 1 23  
 CC FT CHAIN 1 23  
 CC FT DOMAIN 24 630  
 CC FT TRANSMEM 24 234  
 CC FT TRANSMEM 235 258  
 CC FT DOMAIN 259 630  
 CC FT DOMAIN 24 123  
 CC FT DOMAIN 124 228  
 CC FT DISULFID 37 47  
 CC FT DISULFID 76 87  
 CC FT CAROHRD 92 92  
 CC FT CAROHRD 101 101  
 CC FT CAROHRD 101 101  
 CC SQ SEQUENCE 630 AA; 70810 MW; A451563F3D12979D CRC64;

Query Match 54.3%; Score 38; DB 1; Length 630;  
 Best Local Similarity 66.7%; Pred. No. 43;  
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 3 WELEMLPCA 11  
 ||| :|||  
 Db 392 WEKEMPCA 400

RESULT 8  
 GUNC\_CELFI STANDARD; PRT; 1101 AA.  
 AC P14050;  
 DT 01-JAN-1990 (Rel. 13, Created)  
 DT 01-MAR-1992 (Rel. 21, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE ENDOGLUCANASE C PRECURSOR (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE C)  
 DE (CELLULOSE C).  
 GN CENC.  
 OS Cellulomonas fimi.  
 OS Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Micrococcales; Cellulomonadaceae; Cellulomonas.  
 CC [1]  
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 33-42.  
 RC STRAIN-ATCC 484;  
 RX MEDLINE-92065819; PubMed=1956299;  
 RA Coutinho J.B., Moser B., Kilburn D.G., Warren R.A.J., Miller R.C. Jr.;  
 RT "Nucleotide sequence of the endoglucanase C gene (cenc) of  
 RT Cellulomonas fimi, its high-level expression in Escherichia coli, and  
 RT characterization of its products."  
 RL Mol. Microbiol. 5:1221-1233(1991).  
 RN [2]  
 RP SEQUENCE OF 1-64 FROM N.A., AND SEQUENCE OF 625-641.  
 RX MEDLINE-90103465; PubMed=2604391;

RA Moser B., Gilkes N.R., Kilburn D.G., Warren R.A.J., Miller R.C. Jr.;  
 RT "Purification and characterization of endoglucanase C of Cellulomonas  
 RT fimi, cloning of the gene, and analysis of in vivo transcripts of the  
 RT gene."  
 RL Appl. Environ. Microbiol. 55:2480-2487(1989).  
 RN [3]  
 RP CELLULOSE-BINDING DOMAINS.  
 RX MEDLINE-92269585; PubMed=1375311;  
 RA Coutinho J.B., Gilkes N.R., Warren R.A.J., Kilburn D.G.,  
 RA Miller R.C. Jr.;  
 RT "The binding of Cellulomonas fimi endoglucanase C (cenc) to cellulose  
 RT and Sephadex is mediated by the N-terminal repeats."  
 RL Mol. Microbiol. 6:1243-1252(1992).  
 RN [4]  
 RP IDENTIFICATION OF IG-LIKE DOMAINS.  
 RX MEDLINE-97035265; PubMed=8880921;  
 RA Bateman A., Eddy S.R., Chothia C.;  
 RT "Members of the immunoglobulin superfamily in bacteria."  
 RL Protein Sci. 5:1939-1942(1996).  
 RN [5]  
 RP STRUCTURE BY NMR OF 33-184.  
 RX MEDLINE-97074498; PubMed=8916925;  
 RA Johnson P.E., Josh M.D., Tomme P., Kilburn D.G., McIntosh L.P.;  
 RT "Structure of the N-terminal cellulose-binding domain of Cellulomonas  
 RT fimi Cenc determined by nuclear magnetic resonance spectroscopy."  
 RL Biochemistry 35:14381-14394(1996).  
 CC -1- FUNCTION: THE BIOLOGICAL CONVERSION OF CELLULOSE TO GLUCOSE  
 CC GENERALLY REQUIRES THREE TYPES OF HYDROLYTIC ENZYMES:  
 CC (1) ENDOGLUCANASES WHICH CUT INTERNAL BETA-1,4-GLUCOSIDIC BONDS;  
 CC (2) EXOCELLULOBIOSACCHARIDASES THAT CUT THE DISSACCHARIDE CELLULOSE  
 CC FROM THE NONREDUCING END OF THE CELLULOSE POLYMER CHAIN;  
 CC (3) BETA-1,4-GLUCOSIDASES WHICH HYDROLYZE THE CELLOBIOSE AND OTHER  
 CC SHORT CELLO-OLIGOSACCHARIDES TO GLUCOSE.  
 CC -1- CATALYTIC ACTIVITY: ENDOMONOLYSE OF 1,4-BETA-D-GLUCOSIDIC  
 CC LINKAGES IN CELLULOSE.  
 CC -1- SIMILARITY: BELONGS TO CELLULOSE-BINDING DOMAIN E (FAMILY 9 OF GLYCOSYL  
 CC HYDROLASES).  
 CC -1- SIMILARITY: CONTAINS 2 CELLULOSE-BINDING DOMAIN (CBD) REMOTELY  
 CC RELATED TO BACTERIAL-TYPE CBD'S.  
 CC -1- SIMILARITY: CONTAINS 2 IMMUNOGLOBULIN-LIKE DOMAINS.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: X57858; CAA40993.1; -  
 CC EMBL: M29707; AAA23087.1; ALT\_TERM.  
 CC EMBL: M29708; AAA23088.1; ALT\_SEQ.  
 CC PIR: S15271; S15271.  
 CC DR PDB: 1U0L; 01-APR-97.  
 CC DR PDB: 1U1P; 01-APR-97.  
 CC DR INTERPRO: IPR001701; -  
 CC DR INTERPRO: IPR003006; -  
 CC DR PRAM: PR02018; CBD\_6; 2.  
 CC DR PRAM: PR00759; Glyco\_hydro\_9; 1.  
 CC DR PRAM: PR00047; Ig; 1.  
 CC DR PROSITE: PS00592; GLYCOSYL\_HYDROL\_F9\_1; FALISE\_NRG.  
 CC DR PROSITE: PS00698; GLYCOSYL\_HYDROL\_F9\_2; 1.  
 CC KW Cellulose degradation; Hydrolase; Glycosidase; Signal; Repeat;  
 CC Immunoglobulin domain; 3d-structure.  
 CC FT SIGNAL 1 32  
 CC FT CHAIN 1 32  
 CC FT DOMAIN 33 1101  
 CC FT DOMAIN 64 173  
 CC FT DOMAIN 212 318  
 CC FT DOMAIN 329 880  
 CC FT DOMAIN 918 1006  
 CC FT DOMAIN 1008 1097  
 CC FT ACT\_SITE 831 831  
 CC FT ACT\_SITE 882 882  
 CC ENDUGLUCANASE C.  
 CC CELLULOSE-BINDING 1.  
 CC CELLULOSE-BINDING 2.  
 CC CATALYTIC.  
 CC IG-LIKE 1.  
 CC IG-LIKE 2.  
 CC BY SIMILARITY.  
 CC BY SIMILARITY.

FT ACT SITE 891 891 BY SIMILARITY.  
SQ SEQUENCE 1101 AA; 115216 MW; 1FBDAD189CCCF8B5D CRC64;

Query Match  
Best Local Similarity 83.3%; Score 38; DB 1; Length 1101;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 3 WELEWL 8  
DB 560 WELEWM 565

RESULT 9  
UL90\_HCMVA STANDARD; PRT; 66 AA.  
ID UL90\_HCMVA  
AC P16796;  
DT 01-AUG-1990 (Rel. 15, Created)  
DT 01-AUG-1990 (Rel. 15, Last sequence update)  
DT 01-FEB-1991 (Rel. 17, Last annotation update)  
DE HYPOTHETICAL PROTEIN UL90.  
UL90.  
Human cytomegalovirus (strain AD169).  
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
OC Betaherpesvirinae; Cytomegalovirus.

RN SEQUENCE FROM N.A.  
RX MEDLINE-90269039; PubMed-2161319;  
RA Chee M.S., Bankier A.T., Beck S., Bohni R., Brown C.M., Cerny R.,  
Horsnell T., Hutchinson C.A. III, Kouzarides T., Martignetti J.A.,  
RA Preddie E., Satchwell S.C., Tomlinson P., Weston K.M., Barrell B.G.;  
RT "Analysis of the protein-coding content of the sequence of human  
cytomegalovirus strain AD169."  
RL Curr. Top. Microbiol. Immunol. 154:125-169(1990).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; X17403; CAA35364.1; -  
DR PIR; S09855; S09855.  
KW Hypothetical protein.  
SQ SEQUENCE 66 AA; 7445 MW; 27A867762FD4DBCO CRC64;

Query Match  
Best Local Similarity 62.5%; Score 37; DB 1; Length 66;  
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 3 WELEWLPC 10  
DB 45 WRLTWLTC 52

RESULT 10  
ABME\_MOUSE STANDARD; PRT; 229 AA.  
ID ABME\_MOUSE  
AC P51908;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE APOLIPROTEIN B MRNA EDITING PROTEIN (APOBEC-1).  
GN APOBEC1.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=SMALL INTESTINE;

RX MEDLINE-95286585; PubMed-7768898;  
RA Nakamura M., Oka K., Krushkal J., Kobayashi K., Yamamoto M., Li W.H.,  
RA Chan L.;

RT "Alternative mRNA splicing and differential promoter utilization  
RT determine tissue-specific expression of the apolipoprotein B mRNA-  
RT editing protein (ApoBec1) gene in mice. Structure and evolution of  
RT ApoBec1 and related nucleoside/nucleotide deaminases.";  
RL J. Biol. Chem. 270:13042-13056(1995).

CC -1- FUNCTION: RESPONSIBLE FOR THE POSTTRANSCRIPTIONAL EDITING OF  
CC A CAA CODON FOR GUN TO A UAA CODON FOR STOP IN THE APOB MRNA.

CC -1- COFACTOR: ZINC (BY SIMILARITY).

CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).

CC -1- TISSUE SPECIFICITY: EXPRESSED IN THE LIVER AS WELL AS SMALL  
CC INTESTINE.

CC -1- SIMILARITY: BELONGS TO THE CYTIDINE AND DEOXYCYTIDYLATE DEAMINASES  
CC FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; U21951; AAC52211.1; -  
DR EMBL; U21947; AAC52211.1; JOINED.  
DR EMBL; U21948; AAC52211.1; JOINED.  
DR EMBL; U21949; AAC52211.1; JOINED.  
DR EMBL; U21950; AAC52211.1; JOINED.

DR EMBL; U22262; AAC52212.1; -  
DR EMBL; U22263; AAC52213.1; -  
DR EMBL; U22264; AAC52214.1; -  
DR MGD; MGI:103298; APOBEC1.  
DR INTERPRO; IPR002125; -

DR PFM; PF00383; dCMP\_cyt\_deam: 1.  
DR PROSITE; PS00903; CYT\_DCMP\_DEAMINASES; 1.  
KW MRNA processing; Hydroxylase; Zinc.

FT METAL 61 61 ZINC (BY SIMILARITY).  
FT METAL 93 93 ZINC (BY SIMILARITY).  
FT METAL 96 96 ZINC (BY SIMILARITY).  
FT DOMAIN 180 193 LEU-RICH MOTIF.

SQ SEQUENCE 229 AA; 27521 MW; 1CBCE929056ABAD CRC64;

Query Match  
Best Local Similarity 62.5%; Score 37; DB 1; Length 229;  
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 3 WELEWLPC 10  
DB 86 WFLSMSPC 93

RESULT 11  
ABME\_RAT STANDARD; PRT; 229 AA.  
ID ABME\_RAT  
AC P38483;  
DT 01-OCT-1994 (Rel. 30, Created)  
DT 01-OCT-1994 (Rel. 30, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE APOLIPROTEIN B MRNA EDITING PROTEIN (REPR) (APOBEC-1).  
GN APOBEC1.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=SMALL INTESTINE;  
RX MEDLINE-93289362; PubMed-8511591;  
RA Teng B., Burt C.F., Davidson N.O.;  
RT "Molecular cloning of an apolipoprotein B messenger RNA editing  
RT protein.";

```

RL Science 260:1816-1819(1993).
CC -1- FUNCTION: RESPONSIBLE FOR THE POSTTRANSCRIPTIONAL EDITING OF
CC A CAA CODON FOR GIN TO A UAA CODON FOR STOP IN THE APOB MRNA.
CC -1- COFACTOR: ZINC (BY SIMILARITY).
CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC -1- TISSUE SPECIFICITY: EXPRESSED IN THE LIVER AS WELL AS SMALL
CC INTESTINE.
CC -1- SIMILARITY: BELONGS TO THE CYTIDINE AND DEOXYCYTIDYLATE DEAMINASES
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: L07114; AAA17394.1; -.
CC INTERPRO: IPR002125; -.
CC DR PFAM: PF00383; dCMP_cyt_deam; 1.
CC DR PROSITE: PS00903; CYT_DCMP_DEAMINASES; 1.
CC KM MRNA Processing; Hydrolyase; Zinc.
CC FT METAL 61 61 ZINC (BY SIMILARITY).
CC FT METAL 93 93 ZINC (BY SIMILARITY).
CC FT METAL 96 96 ZINC (BY SIMILARITY).
CC FT METAL 180 193 LEU-RICH MOTIF.
CC FT DOMAIN 180 193 LEU-RICH MOTIF.
CC SQ SEQUENCE 229 AA; 27274 MW; 0876644188278983 CRC64;

Query Match 52.9%; Score 37; DB 1; Length 229;
Best Local Similarity 62.5%; Pred. No. 23;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 WELEWLP 10
DB 86 WFLSWSPC 93

RESULT 12
ABME_HUMAN STANDARD; PRT; 236 AA.
AC P41238;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE APOLOPOPROTEIN B MRNA EDITING PROTEIN (HEPR) (APOBEC-1).
OS APOBEC1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=INTESTINE;
RX MEDLINE=9435963; PubMed=8208612;
RA Hadjiagapou C., Giannoni F., Funahashi T., Skarosi S.F.,
RA Davidson N.O.;
RT "Molecular cloning of a human small intestinal apolipoprotein B mRNA
RT editing protein."
RL Nucleic Acids Res. 22:1874-1879(1994).
RN [2]
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
RC TISSUE=SMALL INTESTINE;
RX MEDLINE=9435963; PubMed=8078915;
RA Lau P.P., Zhu H.-J., Baldini A., Charusangavej C., Chan L.;
RT "Dimeric structure of a human apolipoprotein B mRNA editing protein
RT and cloning and chromosomal localization of its gene."
RL Proc. Natl. Acad. Sci. U.S.A. 91:8522-8526(1994).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=PERIPHERAL BLOOD LEUKOCYTES;
RX MEDLINE=98140126; PubMed=9479499;
RA Fujino T., Navaratnam N., Scott J.;

```

```

RT "Human apolipoprotein B RNA editing deaminase gene (APOBEC1).";
RL Genomics 47:266-275(1998).
CC -1- FUNCTION: RESPONSIBLE FOR THE POSTTRANSCRIPTIONAL EDITING OF
CC A CAA CODON FOR GIN TO A UAA CODON FOR STOP IN THE APOB MRNA.
CC -1- COFACTOR: ZINC (BY SIMILARITY).
CC -1- SUBUNIT: HOMODIMER.
CC -1- TISSUE SPECIFICITY: EXPRESSED EXCLUSIVELY IN THE SMALL INTESTINE.
CC -1- SIMILARITY: BELONGS TO THE CYTIDINE AND DEOXYCYTIDYLATE DEAMINASES
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: L25877; AAA86766.1; -.
CC EMBL: L26234; AAA64230.1; -.
CC DR EMBL: AB009426; BAA23882.1; -.
CC DR EMBL: AB009422; BAA23882.1; JOINED.
CC DR EMBL: AB009423; BAA23882.1; JOINED.
CC DR EMBL: AB009424; BAA23882.1; JOINED.
CC DR EMBL: AB009425; BAA23882.1; JOINED.
CC DR PIR: S45253; S45253.
CC DR MIM: 600130; -.
CC DR INTERPRO: IPR002125; -.
CC DR PFAM: PF00383; dCMP_cyt_deam; 1.
CC DR PROSITE: PS00903; CYT_DCMP_DEAMINASES; 1.
CC KM MRNA Processing; Hydrolyase; Zinc.
CC FT METAL 61 61 ZINC (BY SIMILARITY).
CC FT METAL 93 93 ZINC (BY SIMILARITY).
CC FT METAL 96 96 ZINC (BY SIMILARITY).
CC FT METAL 180 193 LEU-RICH MOTIF.
CC FT DOMAIN 180 193 LEU-RICH MOTIF.
CC FT CONFLICT 83 83 S -> T (IN REF. 1).
CC FT CONFLICT 83 83 S -> T (IN REF. 1).
CC SQ SEQUENCE 236 AA; 28173 MW; 42866DEF9FD1A877 CRC64;

Query Match 52.9%; Score 37; DB 1; Length 236;
Best Local Similarity 62.5%; Pred. No. 24;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 WELEWLP 10
DB 86 WFLSWSPC 93

RESULT 13
ABME_RABIT STANDARD; PRT; 236 AA.
AC P47855;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE APOLOPOPROTEIN B MRNA EDITING PROTEIN (APOBEC-1).
OS APOBEC1.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
RN [1]
RP SEQUENCE FROM N.A., AND MUTAGENESIS.
RC STRAIN=NEW ZEALAND WHITE; TISSUE=SMALL INTESTINE;
RX MEDLINE=94342367; PubMed=8065810;
RA Yamanaka S., Poksay K.S., Balestra M.E., Zeng G.-Q., Innerarity T.L.;
RT "Cloning and mutagenesis of the rabbit Apob mRNA editing protein. A
RT zinc motif is essential for catalytic activity, and noncatalytic
RT auxiliary factor(s) of the editing complex are widely distributed."
RL J. Biol. Chem. 269:21725-21734(1994).
CC -1- FUNCTION: RESPONSIBLE FOR THE POSTTRANSCRIPTIONAL EDITING OF
CC A CAA CODON FOR GIN TO A UAA CODON FOR STOP IN THE APOB MRNA.
CC -1- COFACTOR: ZINC.

```

```

CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC -1- TISSUE SPECIFICITY: EXPRESSED EXCLUSIVELY IN THE INTESTINE.
CC -1- SIMILARITY: BELONGS TO THE CYTIDINE AND DEOXYCYTIDYLATE DEAMINASES
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U10695; AAA56718.1; -
DR INTERPRO: IPR002125; -
DR PFAM: PF00383; dcmp_cyt_deam: 1.
DR PROSITE: PS00903; CYT_DCMP_DEAMINASES; 1.
KW mRNA processing; Hydrolase; Zinc.
FT METAL 61 61
FT METAL 93 93
FT METAL 96 96
FT DOMAIN 180 193
FT METAGEN 61 61
FT METAGEN 61 61
FT METAGEN 63 63
FT METAGEN 92 92
FT METAGEN 93 93
FT METAGEN 96 96
SQ SEQUENCE 236 AA; 27719 MW; AB3041CA5102F1F3 CRC64;

Query Match
Best Local Similarity 52.9%; Score 37; DB 1; Length 236;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 WELEWLP 10
Db 86 WFLSWSPC 93

RESULT 14
CO9_ONCMY STANDARD; PRT; 574 AA.
ID CO9_ONCMY
AC P06682;
DT 01-NOV-1988 (Rel. 06, Created)
DT 01-JAN-1991 (Rel. 20, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE COMPLEMENT COMPONENT C9 (FRAGMENT).
GN C9.
NC Oncorhynchus mykiss (Rainbow trout) (Salmo gairdneri).
CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC Actinopterygii; Neopterygii; Teleostei; Euteleostei;
CC Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.
CC [1]
RP SEQUENCE FROM N.A.
RA MEDLINE-88004404; PubMed-2443347;
RA Stanley K.R., Herz J.;
RT "Topological mapping of complement component C9 by recombinant DNA
RT techniques suggests a novel mechanism for its insertion into target
RT membranes";
RL EMBO J. 6:1951-1957(1987).
RN
RP REVISIONS.
RX MEDLINE-90036879; PubMed-2808363;
RA Haefliger J.-A., Tschopp J., Vial N., Jenne D.E.;
RT "Complete primary structure and functional characterization of the
RT sixth component of the human complement system. Identification of the
RT C5b-binding domain in complement C6";
RL J. Biol. Chem. 264:18041-18051(1989).
CC -1- FUNCTION: C9 IS THE FINAL COMPONENT OF THE COMPLEMENT SYSTEM TO
CC BE ADDED IN THE ASSEMBLY OF THE MEMBRANE ATTACK COMPLEX. IT IS
CC ABLE TO ENTER LIPID BILAYERS, FORMING TRANSMEMBRANE CHANNELS.
CC -1- SIMILARITY: TO COMPLEMENT FACTORS C6, C7, C8, AND TO PERFORIN.

```

```

CC -1- SIMILARITY: CONTAINS 2 TYPE-1 TSP REPEATS.
CC -1- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 LDL-RECEPTOR CLASS A DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X05474; CAA29037.1; ALT_SEQ.
DR PIR: B29677; B29677.
DR HSSP: P01130; IATJ.
DR INTERPRO: IPR000864; -
DR INTERPRO: IPR001862; -
DR INTERPRO: IPR002172; -
DR PFAM: PF01823; MACPF_1.
DR PFAM: PF00057; ldl_recept_a; 1.
DR PFAM: PF00090; tsp_1; 2.
DR PROSITE: PS00022; EGF_1; 1.
DR PROSITE: PS01186; EGF_2; FALSE_NEG.
DR PROSITE: PS01209; LDLRA_1; 1.
DR PROSITE: PS50068; LDLRA_2; 1.
DR PROSITE: PS00279; MAC_PERFORIN; 1.
DR PROSITE: PS50092; TSP1_2.
KW Complement pathway; Complement alternate pathway; Glycoprotein;
KW Plasma; Membrane attack complex; Cytolysis; Transmembrane;
KW EGF-like domain; Repeat.
FT NON_TER 1 1
FT REPEAT 18 76
FT DOMAIN 77 117
FT TRANSMEM 284 300
FT TRANSMEM 305 324
FT DOMAIN 475 511
FT REPEAT 528 574
FT DISULFD 79 91
FT DISULFD 86 104
FT DISULFD 98 112
FT DISULFD 350 376
FT DISULFD 481 497
FT DISULFD 484 499
FT DISULFD 501 510
FT CARBOHYD 3 3
FT CARBOHYD 231 231
FT CARBOHYD 261 261
FT CARBOHYD 533 533
SQ SEQUENCE 574 AA; 64033 MW; CB516A9F76824D59 CRC64;

Query Match
Best Local Similarity 52.9%; Score 37; DB 1; Length 574;
Matches 6; Conservative 0; Mismatches 4; Indels 2; Gaps 1;

QY 1 CXWE--LEWLP 10
Db 19 CWSRSEWTPC 30

RESULT 15
NTBE_MOUSE STANDARD; PRT; 614 AA.
ID NTBE_MOUSE
AC P31651;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE SODIUM- AND CHLORIDE-DEPENDENT BETAINE TRANSPORTER (NA+/CL-
DE BETAINE/GABA TRANSPORTER) (SODIUM- AND CHLORIDE-DEPENDENT GABA
DE TRANSPORTER 2) (GAT2).
GN SLC6A12 OR GABT2 OR GAT-2 OR GAT2.
OS Mus musculus (Mouse).

```



```

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-BRAIN;
RX MEDLINE-92388088; PubMed=1517200;
RA Lopez-Corcuera B., Liu Q.-R., Mandiyan S., Nelson H., Nelson N.;
RT "Expression of a mouse brain cDNA encoding novel gamma-aminobutyric
RT acid transporter."
RL J. Biol. Chem. 267:17491-17493(1992).
CC -|- FUNCTION: TRANSPORTS BETAINE AND GABA. MAY HAVE A ROLE IN
CC REGULATION OF GABAERGIC TRANSMISSION IN THE BRAIN THROUGH THE
CC REUPTAKE OF GABA INTO PRESYNAPTIC TERMINALS, AS WELL AS IN OSMOTIC
CC REGULATION (BY SIMILARITY).
CC -|- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -|- TISSUE SPECIFICITY: BRAIN, LIVER AND KIDNEY.
CC -|- SIMILARITY: BELONGS TO THE SODIUM:NEUROTRANSMITTER SYMPORTER
CC FAMILY (SNF).
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M97632; -, NOT_ANNOTATED_CDS.
DR FIR: A43390; A43390.
DR MGD: MGI:95628; GABT2.
DR INTERPRO: IPR000175; -.
DR INTERPRO: IPR002983; -.
DR PFAM: PF00209; SNF: 1.
DR PRINTS: PR00176; NANEUSPORT.
DR PRINTS: PR01198; BETTRANSPORT.
DR PROSITE: PS00610; NA_NEUOTRAN_SYM_1; 1.
DR PROSITE: PS00754; NA_NEUOTRAN_SYM_2; 1.
DR PROSITE: PS50267; NA_NEUOTRAN_SYM_3; 1.
KW Neurotransmitter transport; Transport; Transmembrane; Glycoprotein;
KW Symport; Multigene family.
FT DOMAIN 1 44
FT TRANSMEM 45 65 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 73 92 1 (POTENTIAL).
FT TRANSMEM 117 137 2 (POTENTIAL).
FT DOMAIN 138 210 3 (POTENTIAL).
FT TRANSMEM 211 229 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 238 255 4 (POTENTIAL).
FT TRANSMEM 291 308 5 (POTENTIAL).
FT TRANSMEM 320 341 6 (POTENTIAL).
FT TRANSMEM 374 393 7 (POTENTIAL).
FT TRANSMEM 423 441 8 (POTENTIAL).
FT TRANSMEM 458 478 9 (POTENTIAL).
FT TRANSMEM 499 518 10 (POTENTIAL).
FT TRANSMEM 538 556 11 (POTENTIAL).
FT DOMAIN 557 614 12 (POTENTIAL).
FT DOMAIN 614 171 CYTOPLASMIC (POTENTIAL).
FT CARBOHYD 171 183 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 183 183 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 614 AA; 69613 MW; 9A6B49EA503725B CRC64;
-----
Query Match 52.9%; Score 37; DB 1; Length 614;
Best Local Similarity 62.5%; Pred. No. 61;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
OY 3 WELEWIPC 10
DB 150 WELPWTTC 157

```



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 11, 2001, 12:56:07 ; Search time 13.32 seconds  
(without alignments)  
14.829 Million cell updates/sec

Title: US-09-722-440-8

Perfect score: 70

Sequence: 1 CWMELMPLCA 11

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Archived: 174772 seqs, 17957048 residues

Total number of hits satisfying chosen parameters: 174772

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

Issued\_Patents\_AA: \*  
1: /cgn2\_6/ptodata/2/1aa/5A.COMB.pep: \*  
2: /cgn2\_6/ptodata/2/1aa/5B.COMB.pep: \*  
3: /cgn2\_6/ptodata/2/1aa/6.COMB.pep: \*  
4: /cgn2\_6/ptodata/2/1aa/PTCUS.COMB.pep: \*  
5: /cgn2\_6/ptodata/2/1aa/Backfile1.pep: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	40.5	57.9	595	2	US-08-677-049-11
2	40	57.1	236	1	US-08-158-682A-4
3	39	55.7	204	1	US-08-185-432-8
4	39	55.7	303	1	US-08-185-432-5
5	39	55.7	737	1	US-08-185-432-2
6	39	55.7	737	1	US-08-185-432-4
7	38.5	55.0	2172	1	US-08-611-107-31
8	38.5	55.0	2257	2	US-08-611-107-10
9	38.5	55.0	2257	2	US-08-422-560A-10
10	37	52.9	139	4	US-08-291-299-6
11	37	52.9	139	4	PCT-US95-10579-6
12	37	52.9	229	1	US-08-158-682A-2
13	37	52.9	229	1	US-08-015-203-2
14	37	52.9	229	1	US-08-687-895-5
15	37	52.9	229	1	US-08-816-241-5
16	37	52.9	229	2	US-09-040-482-5
17	37	52.9	229	3	US-09-040-482-5
18	37	52.9	236	3	US-09-128-395-5
19	37	52.9	236	1	US-08-687-895-4
20	37	52.9	236	2	US-08-816-241-4
21	37	52.9	236	2	US-09-040-482-4
22	37	52.9	614	4	US-08-128-395-4
23	37	52.9	614	4	US-08-291-299-7
24	36.5	52.1	993	1	PCT-US95-10579-7
25	36.5	52.1	993	1	US-08-444-792-2
26	36.5	52.1	1039	5	US-08-445-042-2
27	36	51.4	175	1	Patent No. 5196511
28	36	51.4	176	1	US-08-624-125-6
					Sequence 18, Appl

29	36	51.4	316	1	US-08-260-202A-18	Sequence 18, Appl
30	36	51.4	316	1	US-08-017-114-18	Sequence 18, Appl
31	36	51.4	316	3	US-08-505-307-18	Sequence 18, Appl
32	36	51.4	316	4	PCT-US94-02034-18	Sequence 18, Appl
33	35	50.0	116	1	US-08-687-895-3	Sequence 3, Appl1
34	35	50.0	116	1	US-08-816-241-3	Sequence 3, Appl1
35	35	50.0	116	2	US-09-040-482-3	Sequence 3, Appl1
36	35	50.0	116	3	US-09-128-395-3	Sequence 3, Appl1
37	35	50.0	252	1	US-08-411-777-8	Sequence 8, Appl1
38	35	50.0	252	3	US-09-057-088-8	Sequence 8, Appl1
39	35	50.0	1091	3	US-08-986-485-5	Sequence 8, Appl1
40	35	50.0	1101	3	US-08-986-485-2	Sequence 4, Appl1
41	34.5	49.3	125	3	US-08-959-382-4	Sequence 2, Appl1
42	34.5	49.3	655	3	US-08-959-382-2	Sequence 2, Appl1
43	34	48.6	106	2	US-08-785-065-5	Sequence 11, Appl
44	34	48.6	106	2	US-08-785-065-11	Sequence 6, Appl1
45	34	48.6	125	1	US-08-331-657-6	

#### ALIGNMENTS

RESULT 1  
US-08-677-049-11  
Sequence 11, Application US/08677049  
Patent No. 5856707  
GENERAL INFORMATION:  
APPLICANT: Guimaraes, M. Jorge  
APPLICANT: Bazan, J. Fernando  
APPLICANT: McClanahan, Terrill K.  
APPLICANT: Zlotnick, Albert  
TITLE OF INVENTION: PURIFIED MAMMALIAN NUCLEOBASE PERMEASES;  
TITLE OF INVENTION: NUCLEIC ACIDS; ANTIBODIES  
NUMBER OF SEQUENCES: 12  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: DNAX Research Institute  
STREET: 901 California Avenue  
City: Palo Alto  
STATE: California  
COUNTRY: USA  
ZIP: 94304-1104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/677,049  
FILING DATE: 03-JUL-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/000,788  
FILING DATE: 03-JUL-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Ching, Edwin P.  
REGISTRATION NUMBER: 34,090  
REFERENCE/DOCKET NUMBER: DX0511  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-852-9196  
TELEFAX: 415-496-1200  
INFORMATION FOR SEQ ID NO: 11:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 595 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
FEATURE:  
NAME/KEY: Region  
LOCATION: 235..262  
OTHER INFORMATION: /note= "Encompasses TM 4 of Figure  
OTHER INFORMATION: 4"

NAME/KEY: Region  
LOCATION: 459..521  
OTHER INFORMATION: /note= "Encompasses TM 9 and TM 10"  
OTHER INFORMATION: of Figure 4"  
US-08-677-049-11

Query Match 57.9%; Score 40.5; DB 2; Length 595;  
Best Local Similarity 46.7%; Pred. No. 69;  
Matches 7; Conservative 0; Mismatches 3; Indels 5; Gaps 1;

Qy 1 CXWELEW-----LPC 10  
Db 97 CCMRLSWFWSMRLPC 111

RESULT 2  
US-08-158-682A-4  
Sequence 4, Application US/08158682A  
Patent No. 5434058

GENERAL INFORMATION:

APPLICANT: Davidson, Nicholas O.

TITLE OF INVENTION: Apolipoprotein B RNA Editing Protein:

NUMBER OF SEQUENCES: 18

CORRESPONDENCE ADDRESS:

ADDRESSEE: ARNOLD, WHITE & DURKEE

STREET: 321 No. 5434058th Clark Street, Suite 800

CITY: Chicago

STATE: Illinois

COUNTRY: USA

ZIP: 60610

COMPUTER READABLE FORM:

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/158,682A

FILING DATE:

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Coolley, Ronald B.

REGISTRATION NUMBER: 27,187

REFERENCE/DOCKET NUMBER: ARCD:085

TELECOMMUNICATION INFORMATION:

TELEPHONE: (312) 744-0090

TELEFAX: (312) 245-4961

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 236 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-158-682A-4

Query Match 57.1%; Score 40; DB 1; Length 236;  
Best Local Similarity 62.5%; Pred. No. 31;  
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 3 WELEWLP 10  
Db 86 WSLMSMPC 93

RESULT 3

US-08-185-432-8

Sequence 8, Application US/08185432

Patent No. 5750652

GENERAL INFORMATION:

APPLICANT: Attavanis-Tsakonas, Spyridon

APPLICANT: Busseau, Isabelle

APPLICANT: Diederich, Robert J.  
APPLICANT: Xu, Tian  
APPLICANT: Matsuno, Kenji  
TITLE OF INVENTION: DETEX PROTEINS, NUCLEIC ACIDS, AND  
TITLE OF INVENTION: ANTIBODIES, AND RELATED METHODS AND COMPOSITIONS  
NUMBER OF SEQUENCES: 23  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: PENNIE & EDMONDS  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036-2711

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/185,432  
FILING DATE: 21-JAN-1994

CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: Mistrock, S. Leslie  
REGISTRATION NUMBER: 18,872  
REFERENCE/DOCKET NUMBER: 7326-006  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-8864/9741

INFORMATION FOR SEQ ID NO: 8:

SEQUENCE CHARACTERISTICS:

LENGTH: 204 amino acids

TYPE: amino acid

TOPOLOGY: unknown

MOLECULE TYPE: protein

US-08-185-432-8

Query Match 55.7%; Score 39; DB 1; Length 204;  
Best Local Similarity 53.8%; Pred. No. 38;  
Matches 7; Conservative 0; Mismatches 2; Indels 4; Gaps 1;

Qy 1 CXWELEW---WLP 9  
Db 47 CWMFERSGKWL 59

RESULT 4

US-08-185-432-5

Sequence 5, Application US/08185432

Patent No. 5750652

GENERAL INFORMATION:

APPLICANT: Attavanis-Tsakonas, Spyridon

APPLICANT: Busseau, Isabelle

APPLICANT: Diederich, Robert J.

APPLICANT: Xu, Tian

APPLICANT: Matsuno, Kenji

TITLE OF INVENTION: DETEX PROTEINS, NUCLEIC ACIDS, AND

TITLE OF INVENTION: ANTIBODIES, AND RELATED METHODS AND COMPOSITIONS

NUMBER OF SEQUENCES: 23

CORRESPONDENCE ADDRESS:

ADDRESSEE: PENNIE & EDMONDS

STREET: 1155 Avenue of the Americas

CITY: New York

STATE: New York

COUNTRY: U.S.A.

ZIP: 10036-2711

COMPUTER READABLE FORM:

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/185,432  
FILING DATE: 21-JAN-1994  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: Mistrock, S. Leslie  
REGISTRATION NUMBER: 18,872  
REFERENCE/DOCKET NUMBER: 7326-006  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-8864/9741  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 303 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
US-08-185-432-5

Query Match 55.7%; Score 39; DB 1; Length 303;  
Best Local Similarity 53.8%; Pred. No. 58;  
Matches 7; Conservative 0; Mismatches 2; Indels 4; Gaps 1;

QY 1 CXWLE---WLP 9  
| | | | |  
DB 47 CWFEESRGRWLP 59

RESULT 5  
US-08-185-432-2  
Sequence 2, Application US/08185432  
Patent No. 5750652  
GENERAL INFORMATION:  
APPLICANT: Artavanis-Tsakonas, Spyridon  
APPLICANT: Bussseau, Isabelle  
APPLICANT: Diederich, Robert J.  
APPLICANT: Xu, Tian  
APPLICANT: Matsuno, Kenji  
TITLE OF INVENTION: DELTEX PROTEINS, NUCLEIC ACIDS, AND  
TITLE OF INVENTION: ANTIBODIES, AND RELATED METHODS AND COMPOSITIONS  
NUMBER OF SEQUENCES: 23  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: PENNIE & EDMONDS  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036-2711  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/185,432  
FILING DATE: 21-JAN-1994  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: Mistrock, S. Leslie  
REGISTRATION NUMBER: 18,872  
REFERENCE/DOCKET NUMBER: 7326-006  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-8864/9741  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 737 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: protein

US-08-185-432-2

Query Match 55.7%; Score 39; DB 1; Length 737;  
Best Local Similarity 53.8%; Pred. No. 1.5e+02;  
Matches 7; Conservative 0; Mismatches 2; Indels 4; Gaps 1;

QY 1 CXWLE---WLP 9  
| | | | |  
DB 47 CWFEESRGRWLP 59

RESULT 6  
US-08-185-432-4  
Sequence 4, Application US/08185432  
Patent No. 5750652  
GENERAL INFORMATION:  
APPLICANT: Artavanis-Tsakonas, Spyridon  
APPLICANT: Bussseau, Isabelle  
APPLICANT: Diederich, Robert J.  
APPLICANT: Xu, Tian  
APPLICANT: Matsuno, Kenji  
TITLE OF INVENTION: DELTEX PROTEINS, NUCLEIC ACIDS, AND  
TITLE OF INVENTION: ANTIBODIES, AND RELATED METHODS AND COMPOSITIONS  
NUMBER OF SEQUENCES: 23  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: PENNIE & EDMONDS  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036-2711  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/185,432  
FILING DATE: 21-JAN-1994  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: Mistrock, S. Leslie  
REGISTRATION NUMBER: 18,872  
REFERENCE/DOCKET NUMBER: 7326-006  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-8864/9741  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 737 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
US-08-185-432-4

Query Match 55.7%; Score 39; DB 1; Length 737;  
Best Local Similarity 53.8%; Pred. No. 1.5e+02;  
Matches 7; Conservative 0; Mismatches 2; Indels 4; Gaps 1;

QY 1 CXWLE---WLP 9  
| | | | |  
DB 47 CWFEESRGRWLP 59

RESULT 7  
US-08-611-107-31  
Sequence 31, Application US/08611107  
Patent No. 5801233  
GENERAL INFORMATION:  
APPLICANT: Haselkorn, Robert

APPLICANT: Gornicki, Piotr  
TITLE OF INVENTION: NUCLEIC ACID COMPOSITIONS ENCODING  
TITLE OF INVENTION: ACETYL-COA CARBOXYLASE AND USES  
TITLE OF INVENTION: THEREFOR  
NUMBER OF SEQUENCES: 40  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Arnold, White & Durkee  
STREET: P.O. Box 4433  
CITY: Houston  
STATE: Texas  
COUNTRY: United States of America  
ZIP: 77210  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentln Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/611,107  
FILING DATE: Concurrently Herewith  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US SN 07/956,700  
FILING DATE: 02-OCT-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US PCT/US93/09340  
FILING DATE: 30-SEP-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US SN 08/422,560  
FILING DATE: 14-APR-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Kitchell, Barbara S.  
REGISTRATION NUMBER: 33,928  
REFERENCE/DOCKET NUMBER: ARCD:221  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (512) 418-3000  
TELEFAX: (512) 474-7577  
INFORMATION FOR SEQ ID NO: 31:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2172 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
US-08-611-107-31

Query Match 55.0%; Score 38.5; DB 1; Length 2172;  
Best Local Similarity 54.5%; Pred. No. 5.4e+02;  
Matches 6; Conservative 2; Mismatches 2; Indels 1; Gaps 1;  
1 CXMLE-WLPC 10  
1380 COMEYKWLDC 1390

RESULT 8  
US-08-611-107-10  
Sequence 10, Application US/08611107  
Patent No. 5801233  
GENERAL INFORMATION:  
APPLICANT: Haselkorn, Robert  
APPLICANT: Gornicki, Piotr  
TITLE OF INVENTION: NUCLEIC ACID COMPOSITIONS ENCODING  
TITLE OF INVENTION: ACETYL-COA CARBOXYLASE AND USES  
NUMBER OF SEQUENCES: 40  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Arnold, White & Durkee  
STREET: P.O. Box 4433  
CITY: Houston  
STATE: Texas  
COUNTRY: United States of America  
ZIP: 77210

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentln Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/611,107  
FILING DATE: Concurrently Herewith  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US SN 07/956,700  
FILING DATE: 02-OCT-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US PCT/US93/09340  
FILING DATE: 30-SEP-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US SN 08/422,560  
FILING DATE: 14-APR-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Kitchell, Barbara S.  
REGISTRATION NUMBER: 33,928  
REFERENCE/DOCKET NUMBER: ARCD:221  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (512) 418-3000  
TELEFAX: (512) 474-7577  
INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2257 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
US-08-611-107-10

Query Match 55.0%; Score 38.5; DB 1; Length 2257;  
Best Local Similarity 54.5%; Pred. No. 5.4e+02;  
Matches 6; Conservative 2; Mismatches 2; Indels 1; Gaps 1;  
1 CXMLE-WLPC 10  
1420 COMEYKWLDC 1430

RESULT 9  
US-08-422-560A-10  
Sequence 10, Application US/08422560A  
Patent No. 5910626  
GENERAL INFORMATION:  
APPLICANT: Haselkorn, Robert  
APPLICANT: Gornicki, Piotr  
TITLE OF INVENTION: ACETYL-COA CARBOXYLASE COMPOSITIONS AND  
TITLE OF INVENTION: METHODS FOR USE  
NUMBER OF SEQUENCES: 31  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Arnold, White & Durkee  
STREET: P.O. Box 4433  
CITY: Houston  
STATE: TX  
COUNTRY: USA  
ZIP: 77210-4433  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentln Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/422,560A  
FILING DATE: 14-APR-1995  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/956,700  
FILING DATE: 02-OCT-1992  
ATTORNEY/AGENT INFORMATION:



FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: COOLLEY, Ronald B.  
REGISTRATION NUMBER: 27,187  
REFERENCE/DOCKET NUMBER: ARCD:085  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (312) 744-0090  
TELEFAX: (312) 245-4961  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 229 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-158-682A-2

Query Match 52.9%; Score 37; DB 1; Length 229;  
Best Local Similarity 62.5%; Pred. No. 86;  
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 WELEWIPC 10  
| | | | |  
Db 86 WFLSWSPC 93

RESULT 13  
US-08-015-203-2  
Sequence 2, Application US/08015203  
Patent No. 5550034  
GENERAL INFORMATION:  
APPLICANT: Teng, Babie  
APPLICANT: Davidson, Nicholas O.  
APPLICANT: Burtant, Charles F.  
TITLE OF INVENTION: Apolipoprotein B RNA Editing Protein:  
TITLE OF INVENTION: Composition and Method  
NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: ARNOLD, WHITE & DURKEE  
STREET: 321 No. 5550034th Clark Street, Suite 800  
CITY: Chicago  
STATE: Illinois  
COUNTRY: USA  
ZIP: 60610  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/015,203  
FILING DATE: 19930209  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: COOLLEY, Ronald B.  
REGISTRATION NUMBER: 27,187  
REFERENCE/DOCKET NUMBER: ARCD:069  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (312) 744-0090  
TELEFAX: (312) 245-4961  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 229 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-015-203-2

Query Match 52.9%; Score 37; DB 1; Length 229;

Best Local Similarity 62.5%; Pred. No. 86;  
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 WELEWIPC 10  
| | | | |  
Db 86 WFLSWSPC 93

RESULT 14  
US-08-687-895-5  
Sequence 5, Application US/08687895  
Patent No. 5747319  
GENERAL INFORMATION:  
APPLICANT: Au-Young, Janice  
APPLICANT: Hawkins, Phillip R.  
APPLICANT: Hillman, Jennifer L.  
TITLE OF INVENTION: A NOVEL HUMAN MRNA EDITING ENZYME  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: U.S.  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ Version 1.5  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/687,895  
FILING DATE: Filed Herewith  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0109 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 229 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
IMMEDIATE SOURCE:  
LIBRARY: Genbank  
CLONE: 585813  
US-08-687-895-5

Query Match 52.9%; Score 37; DB 1; Length 229;  
Best Local Similarity 62.5%; Pred. No. 86;  
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 WELEWIPC 10  
| | | | |  
Db 86 WFLSWSPC 93

RESULT 15  
US-08-816-241-5  
Sequence 5, Application US/08816241  
Patent No. 5804185  
GENERAL INFORMATION:  
APPLICANT: Bandman, Olga  
APPLICANT: Goli, Surya K.  
TITLE OF INVENTION: NOVEL RNA EDITING ENZYME  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.



STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/816,241  
FILING DATE: Filed Herewith  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0239 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 229 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: GenBank  
CLONE: 585813  
US-08-816-241-5

Query Match 52.9%; Score 37; DB 1; Length 229;  
Best Local Similarity 62.5%; Pred. No. 86;  
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 3 WELEWLPC 10  
1 1 1 1 1  
DB 86 WFLSMSPC 93

Search completed: April 11, 2001, 12:58:16  
Job time: 129 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 11, 2001, 12:56:57 ; Search time 14.05 Seconds

(without alignments)  
53.161 Million cell updates/sec

Title: US-09-722-440-8

Perfect score: 70

Sequence: 1 CXWELEWLPCA 11

Scoring table: BIOSUM62

Gapop 10.0 , Gapext 0.5

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Database : PIR,66:\*

1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	49	70.0	3738	2 T05501	hypothetical prote
2	49	70.0	3795	2 T00831	hypothetical prote
3	42	60.0	763	2 B82677	peptidase Xf1479 l
4	41	58.6	546	2 B32688	beta-galactosidase
5	41	58.6	677	2 A32611	beta-galactosidase
6	40.5	57.9	595	2 A48878	uric acid/xanthine
7	40	57.1	147	2 I46625	rearranged T-cell
8	40	57.1	374	2 S54057	probable membrane
9	39	55.7	273	2 T33619	hypothetical prote
10	39	55.7	419	2 A27539	variant surface gl
11	39	55.7	737	2 S47857	basic protein, cyt
12	38.5	55.0	2257	1 A57710	acetyl-CoA carboxy
13	38	54.3	68	2 D53203	hypothetical prote
14	38	54.3	310	1 K1BERH	thymidine kinase (
15	38	54.3	350	1 K1BERC	thymidine kinase (
16	38	54.3	630	2 I51086	prolactin receptor
17	38	54.3	741	2 A83271	hypothetical prote
18	38	54.3	1101	2 S15271	endoglucanase cenc
19	38	54.3	1245	2 S51255	probable membrane
20	37	52.9	66	2 S09855	hypothetical prote
21	37	52.9	229	2 I59577	apolipoprotein B m
22	37	52.9	229	2 J54269	apolipoprotein B m
23	37	52.9	229	2 I48249	apolipoprotein B m
24	37	52.9	233	2 T20671	apolipoprotein B m
25	37	52.9	236	2 I59323	hypothetical prote
26	37	52.9	236	2 A53853	apolipoprotein B m
27	37	52.9	552	2 E75322	probable trehalase
28	37	52.9	574	2 B96777	complement C9 prec
29	37	52.9	614	2 A43390	gamma-aminobutyric

30	37	52.9	3744	2 S46715	hypothetical prote
31	36.5	52.1	1039	2 A34269	integrin alpha-2b
32	36	51.4	110	2 T16586	hypothetical prote
33	36	51.4	176	2 S52982	beta-carotene hydr
34	36	51.4	205	2 J00223	hypothetical 24k p
35	36	51.4	239	2 J01569	pvs protein 2 - K1
36	36	51.4	283	2 D72378	sugar ABC transport
37	36	51.4	303	2 H72450	proline imino-pept
38	36	51.4	313	2 E82671	align protein - pse
39	36	51.4	316	2 A47064	negative regulator
40	36	51.4	316	2 D83550	conserved hypotet
41	36	51.4	347	2 F75319	hypothetical prote
42	36	51.4	366	2 S46736	probable cis-mucon
43	36	51.4	447	2 T39170	hypothetical prote
44	36	51.4	489	2 B70619	env polyprotein -
45	36	51.4	688	1 VCMWV	

## ALIGNMENTS

RESULT 1  
T05501  
hypothetical protein T19K4.210 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 23-Jul-1999  
C:Accession: T05501  
R:Bevan, M.; Medler, H.; Wambutt, R.; Hohseisel, J.; Mewes, H.W.; Mayer, K.F.X.; Schue  
submitted to the Protein Sequence Database, April 1998  
A:Reference number: Z15418  
A:Accession: T05501  
A:Molecule type: DNA  
A:Residues: 1-3738 <BEV>  
A:Cross-references: EMBL:AL022373  
A:Experimental source: cultivar Columbia; BAC clone T19K4  
C:Genetics:  
A:Map position: 4  
A:Introns: 20/1: 114/2: 243/3: 289/1: 333/2: 391/3: 412/3: 432/3: 438/3: 460/3: 482/1  
68/3: 2152/3: 2168/3: 2298/2: 2446/3: 2879/3: 2966/3: 3029/3: 3049/3: 3183/2: 3352/3;  
A:Note: T19K4.210

Query Match 70.0%; Score 49; DB 2; Length 3738;

Best local similarity 63.6%; Pred. No. 8.4;  
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 CXWELEWLPCA 11  
||| ||| ||  
Db 2634 CXWELEWLPCA 2644

RESULT 2  
T00831  
hypothetical protein T13L16.5 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 12-Feb-1999 #sequence\_revision 12-Feb-1999 #text\_change 30-Apr-1999  
C:Accession: T00831  
R:de la Bastie, M.; Hameed, A.; Gnoj, L.; Jensen, K.; Shohdy, N.; Gottesman, T.; Hab  
McComble, W.R.  
submitted to the EMBL Data Library, January 1999  
A:Description: A. thaliana BAC T13L16 from chromosome IV, top arm.  
A:Reference number: Z14205  
A:Accession: T00831  
A:Status: translated from GB/EMBL/DDBI  
A:Molecule type: DNA  
A:Residues: 1-3795 <DEL>  
A:Cross-references: EMBL:AC003952; NID:q2708736; PID:q2708741  
A:Experimental source: cultivar Columbia  
C:Genetics:  
A:Map position: 2  
A:Introns: 20/1: 258/3: 287/2: 323/3: 355/3: 376/3: 396/3: 419/3: 441/1: 458/3: 467/3  
938/3: 3025/3: 3122/3: 3265/3: 3409/3: 3502/3: 3686/3: 3756/3  
A:Note: T13L16.5

Query Match 70.0%; Score 49; DB 2; Length 3795;  
 Best Local Similarity 63.6%; Pred. No. 8.6;  
 Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 CXWELEWLPKA 11  
 |||||  
 Db 2693 CIMEEQLHCA 2703

## RESULT 3

B82677  
 peptidase XP1479 [imported] - Xylella fastidiosa (strain 9a5c)

C:Species: Xylella fastidiosa  
 C>Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 20-Aug-2000  
 C:Accession: B82677  
 R:anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen  
 Nature 406, 151-157, 2000  
 A>Title: The genome sequence of the plant pathogen Xylella fastidiosa.  
 A:Reference number: A82515; MUID:20365717  
 Note: for a complete list of authors see reference number A59328 below

Accession: B82677

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-763 <SIM>

A:Cross-references: GB:AEO03977; GB:AEO03849; NID:g106492; PIDN:AAF84288.1; GSPDB:GN001

A:Experimental source: strain 9a5c

R:Simpon, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A  
 Britones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carret, H  
 as-Neto, E.; Docena, C.; El-Dorri, H.; Facincani, A.P.; Ferreira, A.J.S.  
 submitted to Genbank, June 2000

A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm  
 J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laigt  
 chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E  
 A; Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miranda, E.C.; Miyaki, C.Y.;  
 F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmeri, D.A  
 Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak  
 A; Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir  
 M.; Tshako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z

A:Reference number: A59328

A:Contents: annotation

C:Genetics:

A:Gene: XP1479

Query Match 60.0%; Score 42; DB 2; Length 763;  
 Best Local Similarity 70.0%; Pred. No. 25;  
 Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 CXWELEWLPKA 10  
 |||||  
 Db 23 CIMEEQLHCA 32

## RESULT 4

B32688  
 beta-galactosidase-related protein - human

C:Species: Homo sapiens (man)  
 C>Date: 29-Jun-1990 #sequence\_revision 29-Jun-1990 #text\_change 21-Jul-2000  
 C:Accession: B32688  
 R:Morreau, H.; Galjart, N.J.; Gillemans, N.; Willemsen, R.; van der Horst, G.T.J.; d'Azz  
 J. Biol. Chem. 264, 20655-20663, 1989  
 A>Title: Alternative splicing of beta-galactosidase mRNA generates the classic lysosomal  
 A:Reference number: A32688; MUID:90062209

A:Accession: B32688

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-546 <MOR>

A:Cross-references: NID:M27508; NID:g179420; PIDN:AAA5599.1; PID:g179421

C:Keywords: alternative splicing

Query Match 58.6%; Score 41; DB 2; Length 546;  
 Best Local Similarity 75.0%; Pred. No. 26;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 4 ELEMWLPKA 11  
 |||||  
 Db 489 ELEMWPCS 496

## RESULT 5

A32611  
 beta-galactosidase (EC 3.2.1.23) precursor - human

N:Alternative names: lactase  
 C:Species: Homo sapiens (man)  
 C>Date: 21-May-1990 #sequence\_revision 21-May-1990 #text\_change 29-Sep-1999  
 C:Accession: A32611; B37066; A31673  
 R:Morreau, H.; Galjart, N.J.; Gillemans, N.; Willemsen, R.; van der Horst, G.T.J.; d'  
 J. Biol. Chem. 264, 20655-20663, 1989  
 A>Title: Alternative splicing of beta-galactosidase mRNA generates the classic lysoso  
 A:Reference number: A32688; MUID:90062209

A:Accession: A32688

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-677 <MOR>

A:Cross-references: GB:M27507; NID:g179400; PIDN:AAA51819.1; PID:g179401

R:Yamamoto, Y.; Hake, C.A.; Martin, B.M.; Kretz, K.A.; Aherm-Rindell, A.J.; Naylor, S  
 DNA Cell Biol. 9, 119-127, 1990

A>Title: Isolation, characterization, and mapping of a human acid beta-galactosidase

A:Reference number: A32611; MUID:90262647

A:Accession: A32611

A:Molecule type: mRNA

A:Residues: 1-677 <YAM>

A:Cross-references: EMBL:M34423; NID:g179422; PIDN:AAA51823.1; PID:g179423

A:Experimental source: testis

R:Namba, E.; Suzuki, K. Commun. 173, 141-148, 1990

A>Title: Molecular cloning of mouse acid beta-galactosidase cDNA: sequence, expressio

A:Reference number: A37086; MUID:91076843

A:Accession: B37086

A:Status: preliminary; not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 1-677 <NAN>

R:Oshima, A.; Tsuji, A.; Nagao, Y.; Sakuraba, H.; Suzuki, Y.

Biochem. Biophys. Res. Commun. 157, 238-244, 1988

A>Title: Cloning, sequencing, and expression of cDNA for human beta-galactosidase.

A:Reference number: A31673; MUID:89061717

A:Accession: A31673

A:Molecule type: mRNA

A:Residues: 1-9, 'P', '11-200, 'A', 202-677 <OSH>

A:Cross-references: GB:M22590; NID:g179418; PIDN:AAA51822.1; PID:g179419

A:Experimental source: Placenta

C:Comment: This enzyme is deficient in GM-1 gangliosidosis and Morquio B syndrome.

C:Genetics:

A:Gene: GDB:GLB1

A:Cross-references: GDB:119987; OMIM:230500

A:Map position: 3p21.33-3p21.33

C:Superfamily: beta-galactosidase bga

C:Keywords: alternative splicing; glycoprotein; glycosidase; hydrolase

F:1-23/Domains: signal sequence #status predicted <SIG>

F:24-677/Product: beta-galactosidase #status predicted <MAT>

F:26,247,464,498,542,545,555/Binding site: carbohydrate (Asn) (covalent) #status pred

Query Match 58.6%; Score 41; DB 2; Length 677;  
 Best Local Similarity 75.0%; Pred. No. 32;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 4 ELEMWLPKA 11  
 |||||  
 Db 620 ELEMWPCS 627

## RESULT 6

A48878  
 uric acid/xanthine transport protein - Emericella nidulans  
 N:Alternate names: uric acid/xanthine permease  
 C:Species: Emericella nidulans, Aspergillus nidulans  
 C:Date: 19-May-1995 #sequence\_revision 19-May-1995 #text\_change 04-Sep-1998  
 C:Accession: A48878; S36031  
 R:Gorfinckel, L.; Diallinas, G.; Scazzocchio, C.  
 J. Biol. Chem. 268, 23376-23381, 1993  
 A:Title: Sequence and regulation of the uapA gene encoding a uric acid-xanthine permease  
 A:Reference number: A48878; MUID:94043131  
 A:Accession: A48878  
 A:Molecule type: DNA  
 A:Residues: 1-595 <GOR>  
 A:Cross-references: EMBL:X71807; NID:9298063; PID:9298064  
 C:Genetics:  
 A:Gene: uapA  
 A:Introns: 161/2; 343/3; 554/2  
 C:Superfamily: uric acid/xanthine transport protein  
 Keywords: transmembrane protein

Query Match 57.9%; Score 40.5; DB 2; Length 595;  
 Best Local Similarity 46.7%; Pred. No. 34;  
 Matches 7; Conservative 0; Mismatches 3; Indels 5; Gaps 1;

OY 1 CWLEW-----LPC 10  
 | | | | |  
 DB 97 CCWRLSWFSMLRPC 111

RESULT 7  
 146625  
 rearranged T-cell receptor delta-chain/ Vdelta1.3-delta1.3-delta1.3 - pig (fragment)  
 C:Species: Sus scrofa domestica (domestic pig)  
 C:Date: 21-Feb-1997 #sequence\_revision 21-Feb-1997 #text\_change 23-Jul-1999  
 C:Accession: 146625  
 R:Yang, Y.G.; Ohta, S.; Yamada, S.; Shimizu, M.; Takagaki, Y.  
 J. Immunol. 155, 1981-1993, 1995  
 A:Title: Diversity of T cell receptor delta-chain cDNA in the thymus of a one-month-old  
 A:Reference number: 146623; MUID:95363165  
 A:Accession: 146625  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-147 <YAN>  
 A:Cross-references: GB:DA9566; NID:q1041134; PIDN:BA08510.1; PID:q1041135  
 C:Superfamily: Immunoglobulin V region; immunoglobulin homology  
 Keywords: T-cell receptor

Query Match 57.1%; Score 40; DB 2; Length 147;  
 Best Local Similarity 100.0%; Pred. No. 10;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 WELEWL 8  
 | | | | |  
 DB 114 WELEWL 119

RESULT 8  
 S54057  
 Probable membrane protein YNR075w - yeast (Saccharomyces cerevisiae)  
 N:Alternate names: hypothetical protein N3820  
 C:Species: Saccharomyces cerevisiae  
 C:Date: 08-Jul-1995 #sequence\_revision 01-Sep-1995 #text\_change 20-Jun-2000  
 C:Accession: S54057; S63408; S63413; S63892  
 R:Levesque, H.; Nicaud, J.M.; Lepingle, A.; Gallardin, C.  
 submitted to the EMBL Data Library, May 1995  
 A:Description: Sequence of a 9.2 kb telomeric fragment from the right arm of S. cerevisiae  
 A:Reference number: S54055  
 A:Accession: S54057  
 A:Molecule type: DNA  
 A:Residues: 1-374 <LEV>  
 A:Cross-references: EMBL:X86790; NID:9805055; PIDN:CAA60488.1; PID:9805058

R.Andre, B.; Iraqui Houssaini, I.; Urrestarazu, L.A.; Vissers, S.  
 submitted to the Protein Sequence Database, April 1996  
 A:Reference number: S62920  
 A:Accession: S63408  
 A:Molecule type: DNA  
 A:Residues: 1-374 <AND>  
 A:Cross-references: EMBL:Z71690; NID:q1302613; PIDN:CAA6359.1; PID:q1302614; MIPS:YN  
 A:Experimental source: strain S288C  
 R:Levesque, H.; Lepingle, A.; Nicaud, J.M.; Gallardin, C.  
 submitted to the Protein Sequence Database, April 1996  
 A:Reference number: S63411  
 A:Accession: S63413  
 A:Molecule type: DNA  
 A:Residues: 1-374 <LEW>  
 A:Cross-references: EMBL:Z71690; NID:q1302613; PIDN:CAA6359.1; PID:q1302614; MIPS:YN  
 A:Experimental source: strain S288C  
 R:Levesque, H.; Lepingle, A.; Nicaud, J.M.; Gallardin, C.  
 Yeast 12, 289-295, 1996  
 A:Title: Sequencing of a 9.2 kb telomeric fragment from the right arm of Saccharomyces  
 A:Reference number: S63890; MUID:97060021  
 A:Accession: S63892  
 A:Status: nucleic acid sequence not shown  
 A:Molecule type: DNA  
 A:Residues: 1-374 <LEF>  
 A:Cross-references: EMBL:X86790; NID:9805055; PIDN:CAA60488.1; PID:9805058  
 C:Genetics:  
 A:Gene: SGD:CSO10  
 A:Cross-references: SGD:S0005358; MIPS:YNR075w  
 A:Map position: 14R  
 C:Superfamily: conserved hypothetical protein YKL219w  
 C:Keywords: transmembrane protein  
 F:67-83/Domain: transmembrane #status predicted <TM1>  
 F:244-260/Domain: transmembrane #status predicted <TM2>

Query Match 57.1%; Score 40; DB 2; Length 374;  
 Best Local Similarity 66.7%; Pred. No. 26;  
 Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 3 WELEWLPCA 11  
 | | | | |  
 DB 45 WSLSWLPLA 53

RESULT 9  
 T33619  
 hypothetical protein F40G9.8 - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C:Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 29-Oct-1999  
 C:Accession: T33619  
 R:Graves, T.; Sutterer, C.; Ozerky, P.  
 submitted to the EMBL Data Library, October 1998  
 A:Description: The sequence of C. elegans cosmid F40G9.  
 A:Reference number: Z21378  
 A:Accession: T33619  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-273 <GRA>  
 A:Cross-references: EMBL:AF099919; PIDN:ACG8793.1; GSPDB:GM00021; CESP:F40G9.8  
 C:Experimental source: strain Bristol N2; clone F40G9  
 C:Genetics:  
 A:Gene: CESP:F40G9.8  
 A:Map position: 3  
 A:Introns: 18/3; 107/3; 183/2; 260/1

Query Match 55.7%; Score 39; DB 2; Length 273;  
 Best Local Similarity 71.4%; Pred. No. 28;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 3 WELEWLP 9  
 | | | | |  
 DB 157 WEVSWLP 163

RESULT 10  
A27539  
variant surface glycoprotein YNat 1.1 precursor - Trypanosoma congolense  
C:Species: Trypanosoma congolense  
C>Date: 19-Nov-1988 #sequence\_revision 19-Nov-1988 #text\_change 20-Mar-1998  
C:Accession: A27539  
R:Strickler, J.E.; Binder, D.A.; L'Italien, J.J.; Shumamoto, G.T.; Walt, S.W.; Dalheim, Biochemistry 26, 796-805, 1987  
A:Title: Trypanosoma congolense: structure and molecular organization of the surface glycoprotein  
A:Reference number: A90525; MUID:87185370  
A:Accession: A27539  
A:Molecule type: mRNA  
A:Residues: 1-419 <STR>  
A:Cross-references: GB:M15112; NID:g162432; PID:g162433  
C:Keywords: glycoprotein  
F:1-28/Domain: signal sequence #status predicted <SIG>  
F:29-419/Product: variant surface glycoprotein YNat 1.1 #status predicted <MAT>

Query Match  
Best Local Similarity 55.7%; Score 39; DB 2; Length 419;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 4 ELEMWPCA 11  
1:|||||  
Db 55 EEDWPCA 62

RESULT 11  
S47857  
basic protein, cytosolic - fruit fly (Drosophila melanogaster)  
C:Species: Drosophila melanogaster  
C>Date: 27-Jan-1995 #sequence\_revision 27-Jan-1995 #text\_change 21-Jul-2000  
C:Accession: S47857  
R:Bussseau, I.; Dieckerich, R.J.; Xu, T.; Artavanis-Tsakonas, S. Genetics 136, 585-596, 1994  
A:Title: A member of the Notch group of interacting loci, dextx encodes a cytoplasmic protein  
A:Reference number: S47857; MUID:94200599  
A:Accession: S47857  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-737 <BUS>  
A:Cross-references: EMBL:U09789; NID:g495218; PIDN:AAA18501.1; PID:g495219  
C:Genetics:  
A:Gene: FlyBase:dx  
A:Cross-references: FlyBase:FBgn0000524  
C:Keywords: cytosol

Query Match  
Best Local Similarity 55.7%; Score 39; DB 2; Length 737;  
Matches 7; Conservative 0; Mismatches 2; Indels 4; Gaps 1;

OY 1 CXWLE---WLP 9  
1:|||||  
Db 47 CWEEFESRGKMLP 59

RESULT 12  
A57710  
acetyl-CoA carboxylase (EC 6.4.1.2) - wheat  
C:Species: Triticum aestivum (common wheat)  
C>Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 01-Sep-2000  
C:Accession: A57710  
R:ornicki, P.; Podkowinski, J.; Scappino, L.A.; DiMaio, J.; Ward, E.; Haselkorn, R. Proc. Natl. Acad. Sci. U.S.A. 91, 6860-6864, 1994  
A:Title: Wheat acetyl-Coenzyme A carboxylase: cDNA and protein structure.  
A:Reference number: A57710; MUID:94316597  
A:Accession: A57710  
A:Status: preliminary  
A:Molecule type: mRNA

A:Residues: 1-2257 <GOR>  
A:Cross-references: GB:U0187; NID:g514305; PIDN:AAA19970.1; PID:g514306  
C:Superfamily: human acetyl-CoA carboxylase; biotin carboxylase homology; lipoyl/biotin  
C:Keywords: ligase  
F:29-534/Domain: biotin carboxylase homology <BCH>  
F:661-733/Domain: lipoyl/biotin-binding homology <LPB>  
F:700/Binding site: biotin (lys) (covalent) #status predicted

Query Match  
Best Local Similarity 55.0%; Score 38.5; DB 1; Length 2257;  
Matches 6; Conservative 2; Mismatches 2; Indels 1; Gaps 1;

OY 1 CXWLE---WLP 10  
1:|||||  
Db 1420 COMEVKMLMDC 1430

RESULT 13  
D53203  
hypothetical protein 4 - Desulfovibrio vulgaris (strain Miyazaki)  
C:Species: Desulfovibrio vulgaris  
C>Date: 19-Oct-1995 #sequence\_revision 19-Oct-1995 #text\_change 21-Jul-2000  
C:Accession: D53203  
R:Kitamura, M.; Kojima, S.; Ogasawara, K.; Nakaya, T.; Sagara, T.; Niki, K.; Miura, K. J. Biol. Chem. 269, 5566-5573, 1994  
A:Title: Novel FMN-binding protein from Desulfovibrio vulgaris (Miyazaki F). Cloning  
A:Reference number: A53203; MUID:94164898  
A:Accession: D53203  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-68 <KIT>  
A:Cross-references: GB:D21804; NID:g2970039; PIDN:BA04828.1; PID:g476040

Query Match  
Best Local Similarity 54.3%; Score 38; DB 2; Length 68;  
Matches 6; Conservative 1; Mismatches 4; Indels 6; Gaps 1;

OY 1 CXWLE---WLP 11  
1:|||||  
Db 40 CWRARWRGAGIRPCA 56

RESULT 14  
K1BETH  
thymidine kinase (EC 2.7.1.21) - turkey herpesvirus  
C:Species: turkey herpesvirus  
C>Date: 31-Mar-1990 #sequence\_revision 31-Mar-1990 #text\_change 11-Jun-1999  
C:Accession: A33346  
R:Martin, S.L.; Aparicio, D.I.; Bandyopadhyay, P.K. J. Virol. 63, 2847-2852, 1989  
A:Title: Genetic and biochemical characterization of the thymidine kinase gene from h  
A:Reference number: A33346; MUID:89255069  
A:Accession: A33346  
A:Molecule type: DNA  
A:Residues: 1-310 <MAR>  
A:Cross-references: GB:M26659; NID:g330940; PIDN:AAA46109.1; PID:g330941  
C:Superfamily: herpesvirus thymidine kinase; herpesvirus thymidine kinase homology  
C:Keywords: ATP; DNA biosynthesis; P-loop; phosphotransferase  
F:10-301/Domain: herpesvirus thymidine kinase homology <HTK>  
F:117-24/Region: nucleotide-binding motif A (P-loop)  
F:117-121/Region: nucleotide-binding motif B  
F:23/Binding site: ATP (lys) #status predicted

Query Match  
Best Local Similarity 54.3%; Score 38; DB 1; Length 310;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 3 WELEWL 8  
1:|||||  
Db 220 WELEWL 225

## RESULT 15

KIBBFC

thymidine kinase (EC 2.7.1.21) - turkey herpesvirus (strain Fc-126)

C:Species: turkey herpesvirus

C:Date: 31-Dec-1991 #sequence\_revision 31-Dec-1991 #text\_change 28-Feb-1997

C:Accession: A33375

R:Scott, S.D.; Ross, N.L.J.; Bins, M.M.

J. Gen. Virol. 70, 3055-3065, 1989

A:Title: Nucleotide and predicted amino acid sequences of the Marek's disease virus and

A:Reference number: A33375; MUID:90063552

A:Accession: A33375

A:Molecule type: DNA

A:Residues: 1-350 &lt;SC0&gt;

A:Cross-references: EMBL:D00561

C:Superfamily: herpesvirus thymidine kinase; herpesvirus thymidine kinase homology

C:Keywords: ATP; DNA biosynthesis; P-loop; phosphotransferase

C:10-301/Domain: herpesvirus thymidine kinase homology &lt;HTK&gt;

C:17-24/Region: nucleotide-binding motif A (P-loop)

C:117-121/Region: nucleotide-binding motif B

F:23/Binding site: ATP (lys) #status predicted

Query Match 54.3%; Score 38; DB 1; Length 350;  
Best Local Similarity 83.3%; Pred. No. 52;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 WELEWL 8  
11:111  
DB 220 WELEWL 225

Search completed: April 11, 2001, 12:58:36  
Job time: 99 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 11, 2001, 12:57:42 ; Search time 20.44 Seconds  
(without alignments)  
63.077 Million cell updates/sec

Title: US-09-722-440-8

Perfect score: 70

Sequence: 1 CXWLEWLPCA 11

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Archived: 374700 seqs, 117207915 residues

Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: SP-archaea:\*  
2: SP-bacteria:\*  
3: SP-fungi:\*  
4: SP-human:\*  
5: SP-invertebrate:\*  
6: SP-mammal:\*  
7: SP-mhc:\*  
8: SP-organelle:\*  
9: SP-phage:\*  
10: SP-plant:\*  
11: SP-rodent:\*  
12: SP-virus:\*  
13: SP-invertebrate:\*  
14: SP-unclassified:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	49	70.0	2089	10	09M067 arabidopsis
2	49	70.0	3738	10	065645 arabidopsis
3	49	70.0	3795	10	09SL52 arabidopsis
4	45.5	65.0	677	11	035936 mus musculu
5	43	61.4	958	10	09SH10 Q9sh10 arabidopsis
6	42	60.0	591	5	09VWT6 Q9vwt6 drosophila
7	42	60.0	763	2	09PD00 Q9pd00 xyella fas
8	39	55.7	273	5	09RT29 Q9rt29 caenorhabdi
9	39	55.7	276	2	045986 Q45986 caulobacter
10	39	55.7	737	2	023985 Q23985 drosophila
11	39	55.7	738	5	09W321 Q9w321 drosophila
12	39	55.7	784	5	09VJ64 Q9vj64 drosophila
13	39	55.7	816	5	027537 Q27537 caenorhabdi
14	39	55.7	836	5	09VB05 Q9vb05 drosophila
15	39	55.7	1078	5	09NDY3 Q9ndy3 leishmania
16	38.5	55.0	2257	10	041511 Q41511 triticum ae
17	38.5	55.0	2260	10	041525 Q41525 triticum ae
18	38	54.3	68	2	046605 Q46605 desulfovibr
19	38	54.3	551	4	015389 Q15389 homo sapien

20	38	54.3	1245	3	Q03868 Q03868 saccharomyc
21	38	54.3	1601	2	P71140 P71140 clostridium
22	37.5	53.6	312	11	09Z2C9 09z2c9 mus musculu
23	37.5	53.6	391	4	09Y216 09y216 homo sapien
24	37.5	53.6	401	4	09Y3N3 09y3n3 homo sapien
25	37.5	53.6	470	4	09NXP6 09nxp6 homo sapien
26	37	52.9	132	4	09UE64 09ue64 homo sapien
27	37	52.9	151	2	09L5F6 09l5f6 salmonella
28	37	52.9	154	12	072779 072779 equine infe
29	37	52.9	204	11	09LTP4 09ltp4 mus musculu
30	37	52.9	233	5	P90824 P90824 caenorhabdi
31	37	52.9	235	6	09TU17 09tu17 monodelphis
32	37	52.9	236	4	09UM71 09um71 homo sapien
33	37	52.9	276	10	09SNS1 09sns1 oryza sativ
34	37	52.9	328	11	09SLG9 09slg9 rattus norv
35	37	52.9	352	2	09RST7 09rst7 delnoccous
36	37	52.9	574	4	043417 043417 homo sapien
37	37	52.9	754	10	004332 004332 arabidopsis
38	37	52.9	757	10	09SR14 09sr14 arabidopsis
39	37	52.9	3435	5	09V9E9 09v9e9 drosophila
40	36	51.4	59	5	09W595 09w595 drosophila
41	36	51.4	75	11	061883 061883 mus musculu
42	36	51.4	110	5	021378 021378 caenorhabdi
43	36	51.4	143	11	09QZ26 09qz26 mus musculu
44	36	51.4	205	2	056241 056241 thermus aqu
45	36	51.4	238	11	09JTY9 09jty9 mus musculu

## ALIGNMENTS

### RESULT 1

09M067 ID 09M067 PRELIMINARY: PRT: 2089 AA.  
AC Q9M067;  
DT 01-OCT-2000 (TREMblrel. 15, Created)  
DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)  
DT 01-OCT-2000 (TREMblrel. 15, Last annotation update)  
DE ATW-LIKE PROTEIN (FRAGMENT).  
GN ATG36080.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota: Viridiplantae: Embryophyta: Tracheophyta: Spermatophyta;  
OC Magnoliophyta: eudicotyledons: core eudicots: Rosidae: eurosids II;  
OC Brassicales: Brassicaceae: Arabidopsis.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Wedler H., Wambutt R., Mewes H.W., Lemcke K., Mayer K.F.X.;  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AL161588; CAB81517.1; -;  
FT NON-TER  
SQ SEQUENCE 2089 AA; 236846 MW; 525DAC931BCADA25 CRC64;

Query Match 70.0%; Score 49; DB 10; Length 2089;  
Best Local Similarity 63.6%; Pred. No. 8.2;  
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 CXWLEWLPCA 11  
| | | | | | | | | |  
Db 985 CXWLEWLPCA 995

RESULT 2  
ID 065645 PRELIMINARY: PRT: 3738 AA.  
AC 065645;  
DT 01-AUG-1998 (TREMblrel. 07, Created)  
DT 01-AUG-1998 (TREMblrel. 07, Last sequence update)

DT 01-OCT-2000 (TREMELREL. 15, Last annotation update)  
 DE AM-LIKE PROTEIN.  
 GN 19K4.210.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;  
 OC Brassicales; Brassicaceae; Arabidopsis.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Beyer M., Medler H., Wambutt R., Hobeisel J., Jesse T., Heijnen L.,  
 RA Vos P., Mewes H.W., Mayer K., Schueler C.,  
 RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.  
 DE EMBL; AL022373; CAI8502.1;  
 DR MENDEL; 29831; Arach.3400;29831.  
 DR INTERPRO; IPR000403;  
 DR INTERPRO; IPR003151;  
 DR INTERPRO; IPR003152;  
 DR PFAM; PF00454; P13\_P14\_kinase; 1.  
 DR PFAM; PF02259; FAT; 1.  
 DR PFAM; PF02260; FATC; 1.  
 DR SEQUENCE 3738 AA; 421983 MW; 12C74AB33D3614FE CRC64;

Query Match 70.0%; Score 49; DB 10; Length 3738;  
 Best Local Similarity 63.0%; Pred. No. 15;  
 Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 CXWELMPCPC 11  
 DB 2634 CLMEBQMLHCA 2644

RESULT 3  
 ID 09SL52 PRELIMINARY; PRT; 3795 AA.  
 AC 09SL52;  
 DT 01-MAY-2000 (TREMELREL. 13, Created)  
 DT 01-MAY-2000 (TREMELREL. 13, Last sequence update)  
 DT 01-OCT-2000 (TREMELREL. 15, Last annotation update)  
 DE ATG17930 PROTEIN.  
 GN ATG17930.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;  
 OC Brassicales; Brassicaceae; Arabidopsis.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA STRAIN-CV. COLUMBIA;  
 RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,  
 RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblyum T.V.,  
 RA Buell C.R., Ketchum K.A., Lee J.J., Renning C.M., Koo H., Moffat K.S.,  
 RA Cronin L.A., Shen M., Vanaken S.E., Umayam L., Tallon L.J., Gill J.E.,  
 RA Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,  
 RA Copenhaver G.P., Freuss D., Niernan W.C., White O., Eisen J.A.,  
 RA Salzberg S.L., Fraser C.M., Venter J.C.;  
 RA "Sequence and analysis of chromosome II of Arabidopsis thaliana";  
 RL Nature 402:761-768(1999).  
 DR EMBL; AC006201; AAD20114.2;  
 DR INTERPRO; IPR000403;  
 DR INTERPRO; IPR003151;  
 DR INTERPRO; IPR003152;  
 DR PFAM; PF00454; P13\_P14\_kinase; 1.  
 DR PFAM; PF02259; FAT; 1.  
 DR PFAM; PF02260; FATC; 1.  
 DR SEQUENCE 3795 AA; 429484 MW; 9C5B7005C15CAE1E CRC64;

Query Match 70.0%; Score 49; DB 10; Length 3795;  
 Best Local Similarity 63.0%; Pred. No. 15;  
 Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 CXWELMPCPC 11  
 DB 2693 CLMEBQMLHCA 2703

RESULT 4  
 ID 035936 PRELIMINARY; PRT; 677 AA.  
 AC 035936;  
 DT 01-JAN-1998 (TREMELREL. 05, Created)  
 DT 01-JAN-1998 (TREMELREL. 05, Last sequence update)  
 DT 01-JUN-2000 (TREMELREL. 14, Last annotation update)  
 DE ARACHIDONATE 15-LIPOXYGENASE, TYPE II (EC 1.13.11.33) (8S-LIPOXYGENASE).  
 GN ALOX15B OR ALOX8.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND CHARACTERIZATION.  
 RA MEDLINE=97450967; PubMed=9305900;  
 RA Jisaka M., Kim R.B., Boeglin W.E., Nanney L.B., Brash A.R.;  
 RA "Molecular cloning and functional expression of a phorbol ester-  
 RT inducible 8S-lipoxygenase from mouse skin";  
 RL J. Biol. Chem. 272:24410-24416(1997).  
 RN [2]  
 RP SEQUENCE FROM N.A., AND TISSUE SPECIFICITY.  
 RC STRAIN=NMRI; TISSUE=EPIDERMIS.  
 RX MEDLINE=98186642; PubMed=9518531;  
 RA Krieg P., Kinzig A., Heldt M., Marks F., Fuerstenberger G.;  
 RA "CDNA cloning of a 8-lipoxygenase and a novel epidermis-type  
 RT lipoxygenase from phorbol ester-treated mouse skin";  
 RL Biochim. Biophys. Acta 1391:7-12(1998).  
 CC -1- CATALYTIC ACTIVITY: ARACHIDONATE + O(2) = (5Z,8Z,11Z,13E)-(15S)-  
 CC 15-HYDROPEROXYOOSA-5,8,11,13-TETRAENATE (THE PRODUCT IS RAPIDLY  
 CC CONVERTED TO THE CORRESPONDING 15S-HYDROXY COMPOUND).  
 CC -1- COFACTOR: IRON (BY SIMILARITY).  
 CC -1- PATHWAY: BIOSYNTHESIS OF LEUKOTRIENES.  
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).  
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN EPIDERMIS AND BRAIN. NO  
 CC EXPRESSION FOUND IN HEART, SPLEEN, LIVER, SKELETAL MUSCLE, KIDNEY  
 CC OR TESTIS.  
 CC -1- INDUCTION: BY PHORBOL ESTER.  
 CC -1- SIMILARITY: BELONGS TO THE LIPOXYGENASE FAMILY.  
 DR EMBL; U93277; AAC5356.1;  
 DR EMBL; Y14696; CAA75003.1;  
 DR HSSP; P12530; ILOX.  
 DR MGD; MGI:1098228; Alox15b.  
 DR INTERPRO; IPR000907;  
 DR INTERPRO; IPR001024;  
 DR PFAM; PF00305; lipoxygenase; 1.  
 DR PFAM; PF01477; PLAT; 1.  
 DR PRINTS; PR00087; LIPOXYGENASE.  
 DR PROSITE; PS00081; LIPOXYGENASE\_2; FALSE NEG.  
 DR PROSITE; PS00711; LIPOXYGENASE\_1; FALSE NEG.  
 DR Oxidoreductase; Dioxigenase; Iron; leukotiene biosynthesis;  
 KW Polymorphism  
 FT METAL 374  
 FT METAL 379  
 FT METAL 554  
 FT METAL 677  
 FT VARIANT 32  
 FT VARIANT 32  
 FT VARIANT 38  
 FT VARIANT 38  
 FT VARIANT 58  
 FT VARIANT 58  
 FT VARIANT 76  
 FT VARIANT 76  
 FT VARIANT 413  
 FT VARIANT 413  
 FT VARIANT 536  
 FT VARIANT 536  
 SO SEQUENCE 677 AA; 76230 MW; 78DB1AC9C2F68399 CRC64;

Query Match 65.0%; Score 45.5; DB 11; Length 677;

Best Local Similarity 75.0%; Pred. No. 10;  
Matches 9; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

QY 1 CXWLEWLP 11  
| | | | | | | | | |  
DB 90 CXWLEWLP 101

RESULT 5  
ID 09SH10 PRELIMINARY; PRT: 958 AA.  
AC 09SH10;  
DT 01-MAY-2000 (TREMBlrel. 13, Created)  
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
DE F20D23.9 PROTEIN.  
GN F20D23.9.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
Mangoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;  
OC Brassicales; Brassicaceae; Arabidopsids.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CV. COLUMBIA;  
RA Federici N.A., Palm C.J., Conway A.B., Conn L., Hansen N.F.,  
Altafi H., Araujo R., Huizar L., Rowley D., Buehler E., Dunn P.,  
Gonzalez A., Kremenetskaia I., Kim C., Lenz C., Li J., Liu S.,  
Luros S., Schwartz J., Shinn P., Toriumi M., Vysotskaia V.S.,  
Walker M., Yu G., Ecker J., Theologis A., Davis R.W.;  
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AC007651; AAD5012.1;  
SR SEQUENCE 958 AA; 103089 MW; DF2D71E3371C4026 CRC64;

Query Match 61.4%; Score 43; DB 10; Length 958;  
Best Local Similarity 66.7%; Pred. No. 37;  
Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CXWLEWLP 9  
| | | | | | | | | |  
DB 268 CGMEPRMLP 276

RESULT 6  
ID 09VMT6 PRELIMINARY; PRT: 591 AA.  
AC 09VMT6;  
DT 01-MAY-2000 (TREMBlrel. 13, Created)  
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
DE CG6394 PROTEIN.  
GN CG6394.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BERKELEY;  
RA MEDLINE=20196006; PubMed=10731132;  
RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,  
George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
Brandon R.C., Rogers Y.H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
Beeson K.Y., Benos P.V., Bertan B.P., Bhandari D., Bolshakov S.,  
Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,  
Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,

RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
Dudson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
Hostin D., Houston K.A., Howland T.J., Mei M.-H., Ibegwam C.,  
Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Kethum K.A.,  
Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai X.,  
Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
Merkulov G., Mitsuhashi N.V., Mobarry C., Morris J., Mostrefi A.,  
Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
Nelson D.M., Nelson K.A., Nixon K., Nusskern D.R., Paclet J.M.,  
Palazzo D.R., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,  
Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT "The genome sequence of Drosophila melanogaster."  
RL Science 287:2185-2195(2000).  
DR EMBL: AE003509; AAF4851.1;  
DR FLYBASE: FBgn0030930; CG6394.  
DR INTERPRO: IPR000772;  
DR INTERPRO: IPR001173;  
DR PRAM: PR00535; Glycos.transf.2; 1.  
SR SEQUENCE 591 AA; 68333 MW; 6CC2F7DC38E4CF17 CRC64;

Query Match 60.0%; Score 42; DB 5; Length 591;  
Best Local Similarity 45.5%; Pred. No. 33;  
Matches 5; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 CXWLEWLP 11  
| | | | | | | | | |  
DB 358 CGSIEWVPCS 368

RESULT 7  
ID 09PDA0 PRELIMINARY; PRT: 763 AA.  
AC 09PDA0;  
DT 01-OCT-2000 (TREMBlrel. 15, Created)  
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
DE PEPTIDASE.  
GN XF1479.  
OS Xylella fastidiosa.  
OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;  
OC Xylella.  
OX NCBI\_TaxID=2371;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=9A5C;  
RA MEDLINE=20365717; PubMed=10910347;  
RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,  
Alvarenga R., Alves L.M.C., Araya J.E., Bala G.S., Baptista C.S.,  
Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,  
Bueno M.R.P., Colombo A.A., Camargo L.E.A., Carriero D.M., Carter H.,  
Colauano N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,  
Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorri H.,  
Facinani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,  
Fraga J.C., Franca S.C., Franco M.C., Froime M., Furlan L.R.,  
Garinier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,  
Ho P.L., Hohnselt J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,  
Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,  
Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,

RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,  
 RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,  
 RA Merck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,  
 RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,  
 RA Nhani A.Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,  
 RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,  
 RA Pelxoto B.R., Pereira G.A.G., Pereira H.A.Jr., Pesquero J.B.,  
 RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,  
 RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,  
 RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,  
 RA da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,  
 RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsunako M.H.,  
 RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,  
 RA Zago M.A., Zatz M., Zeldanis J., Zetubal J.C.;  
 RA "The genome sequence of the plant pathogen *Xylella fastidiosa*.";  
 RL Nature 406:151-157(2000).  
 DR EMBL: AE003977; AAF84288.1;  
 DR INTERPRO: IPR001375;  
 DR INTERPRO: IPR002470;  
 DR PRAM: PF00326; Peptidase\_S9; 1.  
 DR PRINTS: PR00862; PROLIGOPTASE.  
 DR SEQUENCE 763 AA; 86186 MW; 9DD060405199EZ4D CRC64;

Query Match 60.0%; Score 42; DB 2; Length 763;  
 Best Local Similarity 70.0%; Pred. No. 43;  
 Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 1 CXMELEWLP 10  
 Db 23 CLIELEWLD 32

RESULT 8  
 O9T279 PRELIMINARY; PRT; 273 AA.  
 AC O9T279;  
 DT 01-MAY-2000 (TREMBlrel. 13, Created)  
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
 DE F40G9.8 PROTEIN.  
 GN F40G9.8.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabdilita; Rhabditoidea;  
 OC Rhabdilitidae; Peloderinae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 MEDLINE=94150718; PubMed=7906398;  
 RA Wilson R., Almscough R., Anderson K., Baynes C., Berks M.,  
 RA Bontfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,  
 RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,  
 RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,  
 RA Jones M., Kershaw J., Kirsten J., Laister N., Latelle P.,  
 RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,  
 RA Parsons J., Percy C., Rifkin L., Koopra A., Saunders D., Showkeen R.,  
 RA Smaison N., Smith A., Sonhammer E., Staden R., Sulston J.,  
 RA Thierly-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,  
 RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.;  
 RA "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
 elegans.";  
 RL Nature 368:32-38(1994).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 RA Graves T., Sultzer C., Ozersky P.;  
 RT "The sequence of C. elegans cosmid F40G9.";  
 RL Submitted (Oct-1998) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 RA Waterston R.;

RL Submitted (Oct-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF099919; AAC68793.1;  
 SQ SEQUENCE 273 AA; 30514 MW; 23F2B3758736E50E CRC64;

Query Match 55.7%; Score 39; DB 5; Length 273;  
 Best Local Similarity 71.4%; Pred. No. 48;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Oy 3 WELEWLP 9  
 Db 157 WEVSWLP 163

RESULT 9  
 O45986 PRELIMINARY; PRT; 276 AA.  
 AC O45986;  
 DT 01-NOV-1996 (TREMBlrel. 01, Created)  
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
 DT 01-NOV-1996 (TREMBlrel. 01, Last annotation update)  
 DE ORF1 (FLAR) (FLAZ), AND ORF2 GENES. COMPLETE CDS (FLAZ).  
 OS Caulobacter crescentus.  
 OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;  
 OC Caulobacter.  
 OX NCBI\_TaxID=76;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CB15;  
 RA Ely B.;  
 RL Submitted (May-1995) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: U27302; AAC62519.1;  
 SQ SEQUENCE 276 AA; 30726 MW; 1D4FC28DBDAE8522 CRC64;

Query Match 55.7%; Score 39; DB 2; Length 276;  
 Best Local Similarity 62.5%; Pred. No. 48;  
 Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 3 WELEWLP 10  
 Db 211 WTCAMWLP 218

RESULT 10  
 O23985 PRELIMINARY; PRT; 737 AA.  
 AC O23985;  
 DT 01-NOV-1996 (TREMBlrel. 01, Created)  
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
 DT 01-JUN-2000 (TREMBlrel. 14, Last annotation update)  
 DE CYTOPLASMIC BASIC PROTEIN.  
 GN DELTEX.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-OREGON R;  
 RA Busseau I., Diederich R.J., Xu T., Artavanis-Tsakonas S.;  
 RT "A member of the Notch group of interacting loci, deltex encodes a  
 cytoplasmic basic protein.";  
 RL Genetics 136:585-596(1994).  
 DR EMBL: U09789; AAA18501.1;  
 DR FLYBASE: FBgn000524; dx.  
 DR INTERPRO: IPR000425;  
 DR INTERPRO: IPR001841;  
 DR PROSITE: PS00221; MIP; UNKNOWN\_1.  
 SQ SEQUENCE 737 AA; 82137 MW; EDA9F8ACDD7A569 CRC64;

Query Match 55.7%; Score 39; DB 5; Length 737;  
 Best Local Similarity 53.8%; Pred. No. 1.3e+02;  
 Matches 7; Conservative 0; Mismatches 2; Indels 4; Gaps 1;

OY 1 CXMLE-----MLP 9  
 Db 47 CWMEFESCKMPL 59

## RESULT 11

Q9W321 PRELIMINARY; PRT; 738 AA.  
 AC Q9W321;  
 DT 01-MAY-2000 (TREMBlrel. 13, Created)  
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
 DT 01-JUN-2000 (TREMBlrel. 14, Last annotation update)  
 DE DX GENE PRODUCT.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BERKELEY;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotler P.,  
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Laslo P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster."  
 RL Science 287:2185-2195(2000).  
 DR EMBL: AE003437; AAF4170.1;  
 DR FLYBASE: FBgn000524; dx.  
 DR INTERPRO: IPR000425;  
 DR PROSITE: PS00221; MIP; UNKNOWN\_1.  
 SO SEQUENCE 738 AA; 82186 MW; 7B1CF29E024D26AF CRC64;

Query Match 55.7%; Score 39; DB 5; Length 738;  
 Best Local Similarity 53.8%; Pred. No. 1.3e+02;  
 Matches 7; Conservative 0; Mismatches 2; Indels 4; Gaps 1;

OY 1 CXMLE-----MLP 9  
 Db 47 CWMEFESCKMPL 59

## RESULT 12

Q9W321 PRELIMINARY; PRT; 784 AA.  
 AC Q9W321;  
 DT 01-MAY-2000 (TREMBlrel. 13, Created)  
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
 DT 01-JUN-2000 (TREMBlrel. 14, Last annotation update)  
 DE CG10383 PROTEIN.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BERKELEY;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotler P.,  
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Laslo P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster."  
 RL Science 287:2185-2195(2000).  
 DR EMBL: AE003600; AAF3691.1;  
 DR FLYBASE: FBgn0032699; CG10383.  
 SO SEQUENCE 784 AA; 89791 MW; 0819C06773FBE58 CRC64;

Query Match 55.7%; Score 39; DB 5; Length 784;  
 Best Local Similarity 71.4%; Pred. No. 1.4e+02;

Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 Oy 3 MELEMLP 9  
 Db 508 WPEWMLP 514

## RESULT 13

Q27537

ID Q27537

PRELIMINARY;

PRT: 816 AA.

AC Q27537;

DT 01-NOV-1996 (TREMblrel. 01, Created)

DT 01-NOV-1996 (TREMblrel. 01, last sequence update)

DE 01-OCT-2000 (TREMblrel. 15, last annotation update)

GN ZK127.7

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;

OC Rhabditidae; Pelodierinae; Caenorhabditis.

OX NCBI\_TaxID=6239;

[1]

SEQUENCE FROM N.A.

STRAIN-BRISTOL N2;

MEDLINE=94150718; PubMed=7906398;

RA Wilson R., Almscough R., Anderson K., Baynes C., Berks M.,

RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,

RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,

RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,

RA Jones M., Kerhaw J., Kirsten J., Laister N., Latreille P.,

RA Lightning J., Lloyd C., Murray A., Mortimore B., O'Callaghan M.,

RA Parsons J., Percy C., Ritken L., Roopra A., Saunders D., Shownkeen R.,

RA Smailson J., Smith A., Sonhammer E., Staden R., Sulston J.,

RA Thierri-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,

RA Watson A., Waterston L., Wilkinson-Sproat J., Wohlman P.,

RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.

elegans."

RL Nature 368:32-38(1994).

RN [2]

SEQUENCE FROM N.A.

STRAIN-BRISTOL N2;

Du Z.;

[3]

SEQUENCE FROM N.A.

STRAIN-BRISTOL N2;

Waterston R.;

Submitted (MAY-1996) to the EMBL/Genbank/DBJ databases.

[4]

SEQUENCE FROM N.A.

STRAIN-BRISTOL N2;

Waterston R.;

Submitted (APR-1997) to the EMBL/Genbank/DBJ databases.

[5]

SEQUENCE FROM N.A.

STRAIN-BRISTOL N2;

Waterston R.;

Submitted (DEC-1997) to the EMBL/Genbank/DBJ databases.

-1- CATALYTIC ACTIVITY: ATP-DEPENDENT BREAKAGE, PASSAGE AND REJOINING

OF DOUBLE-STRANDED DNA.

EMBL: U58758; AAB93429.1;

DR HSSP: P06786; 1BCW;

DR INTERPRO: IPR001154;

DR INTERPRO: IPR001241;

DR INTERPRO: IPR002205;

DR PFAM: PF00204; DNA\_topoisom; 1.

DR PFAM: PF00521; DNA\_topoisom; 1.

DR PRINTS: PR01158; TOPISMASETI

DR PROSITE: PS00177; TOPISOMERASE\_II; 1.

DR PRODOM: PD000616; -; 1.

KW Isomerase; Topoisomerase; ATP-binding.

SEQUENCE 816 AA; 93029 MW; B3A0AC61D061AEB CRC64;

Query Match 55.7%; Score 39; DB 5; Length 816;  
 Best Local Similarity 50.0%; Pred. No. 1.4e+02;  
 Matches 5; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Oy 1 CXWELEWLP 10

Db 316 CSMNLHSLPC 325

## RESULT 14

Q9VB05

ID Q9VB05

AC Q9VB05;

DT 01-MAY-2000 (TREMblrel. 13, Created)

DT 01-MAY-2000 (TREMblrel. 13, last sequence update)

DE 01-MAY-2000 (TREMblrel. 13, last annotation update)

GN CG12876

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OX Ephydroidea; Drosophilidae; Drosophila.

NCBI\_TaxID=7227;

[1]

SEQUENCE FROM N.A.

STRAIN-BERKELEY;

MEDLINE=20196006; PubMed=10731132;

RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,

RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,

RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,

RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,

RA Burtin R.C., Rogers Y.-H.C., Blazek R.G., Champe M., Pfeiffer B.D.,

RA Wan K.H., Doyle J., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,

RA Abell J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,

RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,

RA Beeson K.T., Benos P.V., Berman B.P., Brokstein P., Brotter P.,

RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,

RA Burtin R.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,

RA Cherry J.M., Cawley S., Dahlke C., Daventport L.B., Davies P.,

RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,

RA Dodson K., Dou P.L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,

RA Durbin J., Evangelista C.C., Ferraz C., Ferrier S., Fleischmann W.,

RA Foster C., Gabriellian A.E., Gary N.S., Gelbart W.M., Glasser K.,

RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,

RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,

RA Hostin D., Houston K.A., Howland T.J., Mei M.-H., Ibegam C.,

RA Jallali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,

RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kuip D., Lai Z.,

RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,

RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,

RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,

RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,

RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,

RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,

RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,

RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,

RA Spraker E., Spradling A.C., Stapleton M., Strong R., Sun E.,

RA Svirskas R., Tector C., Turner C., Venter E., Wang A., Wang X.,

RA Wang Z.-Y., Wasserman D.A., Waterston G.M., Weisenbach J.,

RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,

RA Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,

RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,

RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;

"The genome sequence of Drosophila melanogaster.";

Science 287:2185-2195(2000)

EMBL: AE003762; AAF5740.1;

DR EMBL: F89003941; CG12876.

DR EMBL: F89003941; CG12876.

SEQUENCE 836 AA; 92538 MW; 0E02FEDA6B9CF4DD CRC64;

Query Match 55.7%; Score 39; DB 5; Length 836;

Best Local Similarity 71.4%; Pred. No. 1.4e+02;

Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Fri Apr 20 13:16:34 2001

OY 3 WELEWLP 9  
11111  
DB 239 WELEWLP 245

RESULT 15

O9NDY3 PRELIMINARY; PRT: 1078 AA.

AC O9NDY3.

DT 01-OCT-2000 (TREMBlrel. 15, Created)

DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)

DE HYPOTHETICAL 117.2 KDA PROTEIN.

GN L5856.06.

OS Leishmania major.

OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.

NCBI\_TaxID=5664;

[1]

SEQUENCE FROM N.A.

RC STRAIN-FRIEDLIN;

RA Bothe G., Pohl T., Ivens A.C., Quail M., Rajandream M.A.,

RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.

[2]

SEQUENCE FROM N.A.

RC STRAIN-FRIEDLIN;

RA Ivens A.C., Lewis S.M., Bagherzadeh A., Zhang L., Chan H.M.,

RL Smith D.F.; "A physical map of the Leishmania major Friedlin genome.";

RT Genome Res. 8:135-145(1998).

DR EMBL; AL357592; CAB93468.1; -.

KW Hypothetical protein.

SQ SEQUENCE 1078 AA; 117186 MW; 6E0C4CC64566A120 CRC64;

Query Match 55.7%; Score 39; DB 5; Length 1078;

Best Local Similarity 71.4%; Pred. No. 1.9e+02;

Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 3 WELEWLP 9

11111

DB 191 WELEWLP 197

arch completed: April 11, 2001, 12:59:46

Time: 124 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 11, 2001, 12:55:42 ; Search time 19.49 Seconds  
(without alignments)  
19.299 Million cell updates/sec

Title: US-09-722-440-8  
Perfect score: 70  
Sequence: 1 CXMELEWLPCA 11

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Database: 268485 seqs, 34193795 residues  
Total number of hits satisfying chosen parameters: 268485

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: A.Geneseq.36.\*  
2: /SIDSL/gcgdata/geneseq/geneseqp/AA1980.DAT.\*  
3: /SIDSL/gcgdata/geneseq/geneseqp/AA1981.DAT.\*  
4: /SIDSL/gcgdata/geneseq/geneseqp/AA1982.DAT.\*  
5: /SIDSL/gcgdata/geneseq/geneseqp/AA1983.DAT.\*  
6: /SIDSL/gcgdata/geneseq/geneseqp/AA1984.DAT.\*  
7: /SIDSL/gcgdata/geneseq/geneseqp/AA1985.DAT.\*  
8: /SIDSL/gcgdata/geneseq/geneseqp/AA1986.DAT.\*  
9: /SIDSL/gcgdata/geneseq/geneseqp/AA1987.DAT.\*  
10: /SIDSL/gcgdata/geneseq/geneseqp/AA1988.DAT.\*  
11: /SIDSL/gcgdata/geneseq/geneseqp/AA1989.DAT.\*  
12: /SIDSL/gcgdata/geneseq/geneseqp/AA1990.DAT.\*  
13: /SIDSL/gcgdata/geneseq/geneseqp/AA1991.DAT.\*  
14: /SIDSL/gcgdata/geneseq/geneseqp/AA1992.DAT.\*  
15: /SIDSL/gcgdata/geneseq/geneseqp/AA1993.DAT.\*  
16: /SIDSL/gcgdata/geneseq/geneseqp/AA1994.DAT.\*  
17: /SIDSL/gcgdata/geneseq/geneseqp/AA1995.DAT.\*  
18: /SIDSL/gcgdata/geneseq/geneseqp/AA1996.DAT.\*  
19: /SIDSL/gcgdata/geneseq/geneseqp/AA1997.DAT.\*  
20: /SIDSL/gcgdata/geneseq/geneseqp/AA1999.DAT.\*  
21: /SIDSL/gcgdata/geneseq/geneseqp/AA2000.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	41	58.6	677	19 W70900	Human beta galacto
2	39	55.7	639	20 W81023	Human sialoadhesin
3	39	55.7	737	16 R76640	Deltek protein. H
4	39	55.7	737	16 R76639	Deltek protein. H
5	39	55.7	737	18 W18317	Drosophila Deltek
6	39	55.7	1422	16 R82071	Hepatitis GB virus
7	38.5	55.0	2172	19 W70409	Wheat acetyl-CoA c
8	38.5	55.0	2257	17 W05209	Wheat acetyl-CoA c
9	38.5	55.0	2257	19 W70407	Wheat acetyl-CoA c
10	38.5	55.0	2260	17 W05212	Triticum aestivum
11	38	54.3	187	20 Y60168	Wheat acetyl-CoA c
12	38	54.3	421	18 W25945	Human endometrium
					Ob binding protein

13	38	54.3	551	19 W55884	Human CD33-like pr
14	38	54.3	606	17 R93121	Tilapia prolactin
15	38	54.3	630	17 R93120	Tilapia prolactin
16	37	52.9	15	21 Y91624	Human secreted pro
17	37	52.9	36	21 Y76248	Human secreted pro
18	37	52.9	70	18 W20482	Fragment of human
19	37	52.9	95	18 W20482	H. pylori cytoplas
20	37	52.9	95	18 W20482	H. pylori cytoplas
21	37	52.9	139	17 R89482	Human secreted pro
22	37	52.9	139	17 R89482	Rat beta2-microglob
23	37	52.9	139	17 R89482	Rat beta2-microglob
24	37	52.9	139	17 R89482	Rat beta2-microglob
25	36.5	52.1	1039	21 Y92445	Variant human GPII
26	36.5	52.1	1039	21 Y92445	Variant human GPII
27	36.5	52.1	1039	21 Y92445	Variant human GPII
28	36	51.4	92	20 Y60025	Human endometrium
29	36	51.4	176	12 R13988	Beta-carotene hydr
30	36	51.4	205	15 R54075	Thermal resistance
31	36	51.4	261	19 W71573	Hepatocyte nuclear
32	36	51.4	316	15 R57062	P. aeruginosa mucB
33	36	51.4	316	15 R57062	P. aeruginosa mucB
34	36	51.4	384	20 Y42383	Amino acid sequenc
35	36	51.4	384	21 Y84437	Amino acid sequenc
36	36	51.4	688	20 Y41142	Mouse mammary tumo
37	35	50.0	54	21 Y64796	Human 5' EST relat
38	35	50.0	234	19 W53973	Pseudomonas fluore
39	35	50.0	248	13 R29622	Tet d gene product
40	35	50.0	248	13 R29622	Tet d gene product
41	35	50.0	249	11 R05858	Stem cell leukemia
42	35	50.0	350	19 W62621	Mus musculus SOC57
43	35	50.0	367	14 R34543	Tet d gene product
44	35	50.0	368	13 R20992	Protein "e" encode
45	35	50.0	368	18 W14835	Protein encoded by

## ALIGNMENTS

RESULT 1	
W70900	W70900 standard; Protein; 677 AA.
XX	XX
AC	W70900:
XX	XX
DT	14-OCT-1998 (first entry)
XX	XX
DE	Human beta galactosidase protein.
XX	XX
KW	Retroviral vector; gene delivery vehicle; expression;
KW	non-immunogenic selectable marker; gene therapy; activation;
KW	human; beta galactosidase.
XX	XX
OS	Homo sapiens.
XX	XX
PN	W09830709-A2.
XX	XX
PD	16-JUL-1998.
XX	XX
PF	14-JAN-1998; 98WO-US00715.
XX	XX
PR	13-JAN-1998; 98US-0038339.
PR	14-JAN-1997; 97US-0035473.
PR	27-FEB-1997; 97US-0038339.
XX	XX
PA	(CHAD/) CHADA S.
PA	(JOLLY/) JOLLY D J.
XX	(MOORE/) MOORE M D.
PI	Chada S, Jolly DJ, Moore MD;
XX	XX
DR	WPI: 1998-399153/34.
DR	N-PSDB: V42728.
XX	XX

PT Non-immunogenic pro:drug activating enzyme(s) and selectable  
 PT marker(s) - are used in gene therapy for the treatment of a wide  
 PT variety of disorder(s)  
 XX  
 PS Example 2; Fig 5A-B; 121pp; English.  
 CC The present sequence represents human beta galactosidase and is used as  
 CC a marker in the retroviral vector of the invention. The specification  
 CC describes a gene delivery vehicle which directs expression of a  
 CC non-immunogenic selectable marker or molecule which is capable of  
 CC activating a previously inactive compound. Vectors expressing the  
 CC markers and a heterologous sequence are useful in gene therapy. The  
 CC vectors can be used to deliver a molecule into a target area where it  
 CC may cause the activation of a previously inactive substance.  
 CC  
 SQ Sequence 677 AA;  
 Query Match 58.6%; Score 41; DB 19; Length 677;  
 Best Local Similarity 75.0%; Pred. No. 84;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 4 ELEMPLPCA 11  
 |||||  
 Db 620 elewapcs 627  
 RESULT 2  
 W81023 ID W81023 standard; Protein; 639 AA.  
 XX W81023;  
 AC  
 XX  
 DT 26-APR-1999 (first entry)  
 XX  
 DE Human sialoadhesin family 4 (SAF-1) polypeptide.  
 XX  
 KW SAF-4; sialoadhesin family; human; therapy; diagnosis; cancer;  
 KW inflammation; autoimmune disease; allergy; asthma; inflammation;  
 KW cerebellar degeneration; Alzheimer's disease; Parkinson's disease;  
 KW multiple sclerosis; amyotrophic lateral sclerosis; head injury;  
 KW septic shock; sepsis; stroke; osteoporosis; osteoarthritis;  
 KW ischemia reperfusion injury; cardiovascular disease;  
 KW kidney disease; liver disease; myocardial infarction; hypotension;  
 KW hypertension; AIDS; myelodysplastic syndrome; aplastic anaemia;  
 KW baldness; infection.  
 XX  
 OS Homo sapiens.  
 XX  
 WO9853840-A1.  
 03-DEC-1998.  
 XX  
 PF 27-MAY-1998; 98WO-US10791.  
 XX  
 PR 27-MAY-1997; 97US-0047572.  
 XX  
 PA (SMIK ) SMITHKLINE BEECHAM CORP.  
 XX  
 PI Erickson-Miller CL, Kikly KK;  
 XX  
 DR WPI; 1999-080779/07.  
 DR N-PSDB; V99911.  
 XX  
 PT New sialoadhesin family 4 polypeptides and polynucleotides - useful  
 PT to treat various diseases associated with SAF-4 expression  
 XX  
 PS Claim 1; Page 31; 48pp; English.  
 CC This is the amino acid sequence of new human sialoadhesin family 4  
 CC (SAF-4), as deduced from the nucleotide sequence of an isolated  
 CC cDNA clone (see V99911). SAF-4 polynucleotides and polypeptides,  
 CC and methods for producing such polypeptides in transformed host

CC cells using recombinant techniques are disclosed. SAF-4, its  
 CC agonists and antagonists, and nucleic acid molecules that enhance  
 CC or inhibit SAF-4 expression, may be used to treat patients in need  
 CC of enhancement or inhibition of SAF-4 expression or activity.  
 CC Conditions that may benefit from such treatment include cancer,  
 CC inflammation, autoimmunity, allergy, asthma, Rheumatoid arthritis,  
 CC CNS inflammation, cerebellar degeneration, Alzheimer's disease,  
 CC Parkinson's disease, multiple sclerosis, amyotrophic lateral  
 CC sclerosis, head injury damage and other neurological disorders,  
 CC septic shock, sepsis, stroke, osteoporosis, osteoarthritis,  
 CC ischemia reperfusion injury, cardiovascular disease, kidney  
 CC disease, liver disease, ischemic injury, myocardial infarction,  
 CC hypotension, hypertension, AIDS, myelodysplastic syndromes and  
 CC other hematologic abnormalities, aplastic anaemia, male baldness  
 CC pattern and bacterial, protozoal, fungal and viral infections  
 CC related to SAF-4 polypeptide activity. Methods of identifying  
 CC agonists, antagonists/inhibitors are also provided, as well as  
 CC diagnostic assays for detecting diseases associated with  
 CC inappropriate SAF-4 activity or levels.  
 SQ Sequence 639 AA;  
 Query Match 55.7%; Score 39; DB 20; Length 639;  
 Best Local Similarity 54.5%; Pred. No. 1.6e+02;  
 Matches 6; Conservative 1; Mismatches 4; Indels 0; Gaps 0;  
 1 CXELEMPLPCA 11  
 |||||  
 Db 395 cswaeaglhcs 405  
 RESULT 3  
 R76640 ID R76640 standard; Protein; 737 AA.  
 XX R76640;  
 AC  
 XX  
 DT 10-APR-1996 (first entry)  
 XX  
 DE Deltex protein.  
 XX  
 KW Deltex protein; Cancer; antibody; diagnosis; disease; disorder;  
 KW malignancy; notch; cervical cancer; breast cancer; colon cancer;  
 KW lung cancer; melanoma; seminoma.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO9519779-A1.  
 XX  
 PD 27-JUL-1995.  
 XX  
 PF 20-JAN-1995; 95WO-US00825.  
 XX  
 PR 21-JAN-1994; 94US-0185432.  
 XX  
 PA (UYVA ) UNIV YALE.  
 XX  
 PI Artavanis-tsakonas S, Busseau I, Diederich RJ, Matsuno K;  
 PI Xu T;  
 XX  
 DR WPI; 1995-269267/35.  
 DR N-PSDB; Q92980.  
 XX  
 PT Novel deltex protein and related nucleic acids and antibodies -  
 PT useful in treating/diagnosing a malignancy characterised in aberrant  
 PT level of Notch-deltex protein binding activity, e.g. cancer  
 XX  
 PS Claim 2; Figure 12; 153pp; English.  
 CC Purified deltex protein, nucleic acid encoding it and antibodies  
 CC directed against it can be used in pharmaceutical compositions for  
 CC treating or preventing a disease or disorder where function of a



CC for sequences showing homology to Drosophila Deltex. Such regions  
 CC of homology (see also W18318-25) were used to design primers (see  
 CC also T68925-28) for the amplification of human deltex cDNA (T68923).  
 XX

SO Sequence 737 AA;

Query Match 55.7%; Score 39; DB 18; Length 737;  
 Best Local Similarity 53.8%; Pred. No. 1.8e+02;  
 Matches 7; Conservative 0; Mismatches 2; Indels 4; Gaps 1;  
 OY 1 CXWLE---WLP 9  
 Db 47 cwfefesrgkwp 59

RESULT 6  
 ID R82071 standard; Protein; 1422 AA.  
 XX R82071;  
 XX

02-JUL-1996 (first entry)

XX Hepatitis GB virus (HGBV) clone protein prod.  
 XX  
 XX Hepatitis GB virus; HGBV; diagnosis; treatment; vaccine;  
 KM reagents; non-A; non-B; non-C; non-D; non-E; clone;  
 KM lamartin; infected plasma; lambda phage; cDNA library.  
 XX  
 OS Hepatitis GB virus.

XX Key Location/Qualifiers  
 FH MISC-difference 1..1422  
 FT /note= "others correspond to degenerate or SNP  
 FT codons in T00052"

XX W09521922-A2.

XX 17-AUG-1995.

XX 14-FEB-1995; 95WO-US02118.

XX 27-JAN-1995; 95US-0344557.  
 PR 14-FEB-1994; 94US-0196030.  
 PR 13-MAY-1994; 94US-0242654.  
 PR 29-JUL-1994; 94US-0283314.  
 PR 23-NOV-1994; 94US-0344185.  
 PR 23-NOV-1994; 94US-0344190.

(ABBO) ABBOTT LAB.

PI Buljk SL, Dawson GJ, Desai SM, Erker JC, Leary TP;  
 PI Muerhoff AS, Mushahwar IK, Pilot-Matias TJ, Schlauder GG;  
 PI Simons JN;  
 XX

DR WPI; 1995-293123/38.  
 N-PSDB; T00052.

XX Non-A, non-B, non-C, non-D, non-E Hepatitis virus reagents - useful  
 PT for diagnosis and therapy of hepatitis GB virus  
 XX

XX Example 5; Pages 259-264; 661pp; English.

XX Double stranded hepatitis GB virus (HGBV) DNA obid. from HGBV  
 CC infected tamarin plasma, using standard procedures, was used to  
 CC prepare a lambda phage HGBV cDNA library. The cDNA clone T00052,  
 CC which encodes the proteins R82066-71 (the 6 possible reading  
 CC frames), was rescued from the lambda phage, searched against a  
 CC sequence database and found to be a unique HGBV sequence.  
 CC Reagents which comprise the HGBV DNA, or its protein prods. can  
 CC be used for the diagnosis, therapy or in a vaccine to prevent  
 CC HGBV infection.

XX Sequence 1422 AA;

Query Match 55.7%; Score 39; DB 16; Length 1422;  
 Best Local Similarity 62.5%; Pred. No. 3.6e+02;  
 Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
 OY 3 WELEWLP 10  
 Db 168 wilewmpc 175

RESULT 7  
 ID W70409 standard; Protein; 2172 AA.  
 XX W70409;  
 XX

18-NOV-1998 (first entry)

XX Wheat acetyl-CoA carboxylase as deduced from the genomic DNA sequence.  
 XX  
 XX Triticum aestivum acetyl-CoA carboxylase; ACC; plant oil;  
 KM herbicide resistance.  
 KM  
 XX Triticum aestivum.

XX Key Location/Qualifiers  
 FH MISC-difference 1892  
 FT /note= "Residue not specified in the specification"

XX US5801233-A.

XX 01-SEP-1998.

XX 05-MAR-1996; 96US-0611107.

XX 05-MAR-1996; 96US-0611107.  
 PR 02-OCT-1992; 92US-0956700.  
 PR 14-APR-1995; 95US-0422560.

XX (ARCH-) ARCH DEV CORP.

XX Gornicki P, Haselkorn R;  
 XX

DR WPI; 1998-494841/42.  
 N-PSDB; V33433.

XX Isolated DNA encoding plant and cyanobacterial acetyl-CoA  
 PT carboxylase polypeptides - useful for producing recombinant  
 PT polypeptides and increasing the herbicide resistance of plants  
 XX

PS Claim 2; Columns 119-150; 96pp; English.

XX The present claimed sequence represents the Triticum aestivum  
 CC acetyl-CoA carboxylase (ACC) protein as deduced from the genomic  
 CC DNA (V33433) sequence. The invention claims for novel ACC proteins  
 CC and the DNA encoding these proteins from plant (e.g. wheat and canola)  
 CC and cyanobacterial (e.g. Anabaena and Synchococcus) species. The  
 CC ACCs of the invention are claimed to be useful for regulating the oil  
 CC content of plant tissues, for conferring and modulating herbicide  
 CC resistance in plants, and for altering the activity of ACC in plant  
 CC cells in vivo.

SO Sequence 2172 AA;

Query Match 55.0%; Score 38.5; DB 19; Length 2172;  
 Best Local Similarity 54.5%; Pred. No. 6.5e+02;  
 Matches 6; Conservative 2; Mismatches 2; Indels 1; Gaps 1;

OY 1 CXWLE-WLPC 10

Db 1380 cqwevklwldc 1390

RESULT 8  
W05209  
ID W05209 standard; Protein; 2257 AA.

XX W05209;

DT 06-FEB-1997 (first entry)

XX Wheat acetyl-CoA carboxylase.

XX Acetyl-CoA carboxylase; oilseed; vegetable oil;  
KW arylloxyphenoxypionate; cyclohexanedione; herbicide resistance;  
KW monocot; transgenic plant; crop improvement; wheat.

XX Triticum aestivum var. Tam107, Hard Red Winter.

XX W09632484-A2.

XX 17-OCT-1996.

XX 12-APR-1996; 96WO-US05095.

XX 05-MAR-1996; 96US-0611546.

XX 14-APR-1995; 95US-0422560.

XX 06-JUN-1995; 95US-0468793.

XX (ARCH-) ARCH DEV CORP.

XX Gornicki P, Haselkorn R;

XX WPI; 1996-485466/48.

XX N-PSDB; T43072.

XX DNA encoding plant or cyanobacterial acetyl-CoA carboxylase enzyme  
PT - useful for regulating oil content and for conferring and altering  
PT resistance to herbicides which target acetyl-CoA carboxylase in  
PT monocots

XX Claim 9; Page 147-165; 264pp; English.

XX Wheat acetyl-CoA carboxylase (ACCase) (W05209) catalyses the first  
CC committed step in de novo fatty acid biosynthesis; the addition of  
CC CO2 to acetyl-CoA to yield malonyl-CoA. Its amino acid sequence  
CC was deduced from a cDNA clone (T43072) isolated using ACC-specific  
CC cDNA probes. Plant and cyanobacterial ACCases (W05205-12) can be  
CC used to modulate the oil content and herbicide resistance of  
CC transgenic plants (esp. monocots). They may also be expressed in  
CC transformed host (esp. cyanobacterial, E. coli or monocot plant)  
CC cells and used to raise antibodies useful in the immunodetection of  
CC ACCase.

XX Sequence 2257 AA;

Query Match 55.0%; Score 38.5; DB 17; Length 2257;  
Best Local Similarity 54.5%; Pred. No. 6.7e+02;  
Matches 6; Conservative 2; Mismatches 2; Indels 1; Gaps 1;

QY 1 CXWLE-WLPC 10  
DB 1420 cqwevklwldc 1430

RESULT 9  
W70407  
ID W70407 standard; Protein; 2257 AA.  
XX W70407;  
XX

DT 18-NOV-1998 (first entry)

XX Triticum aestivum acetyl-CoA carboxylase protein.

XX Triticum aestivum acetyl-CoA carboxylase cDNA; ACC; plant oil;  
KW herbicide resistance.

XX Triticum aestivum.

XX US5801233-A.

XX 01-SEP-1998.

XX 05-MAR-1996; 96US-0611107.

XX 05-MAR-1996; 96US-0611107.

XX 02-OCT-1992; 92US-0956700.

XX 14-APR-1995; 95US-0422560.

XX (ARCH-) ARCH DEV CORP.

XX Gornicki P, Haselkorn R;

XX WPI; 1998-494841/42.

XX N-PSDB; V33410.

XX Isolated DNA encoding plant and cyanobacterial acetyl-CoA  
PT carboxylase polypeptides - useful for producing recombinant  
PT polypeptides and increasing the herbicide resistance of plants

XX Claim 5; Columns 89-100; 96pp; English.

XX The present claimed sequence represents the Triticum aestivum  
CC acetyl-CoA carboxylase (ACC) protein which is encoded by the T.  
CC aestivum ACC cDNA (V33410). The invention claims for novel  
CC ACC proteins and the DNA encoding these proteins from plant (e.g.  
CC wheat and canola) and cyanobacterial (e.g. Anabaena and Synechococcus)  
CC species. The ACCs of the invention are claimed to be useful for  
CC regulating the oil content of plant tissues, for conferring and  
CC modulating herbicide resistance in plants, and for altering the  
CC activity of ACC in plant cells in vivo.

XX Sequence 2257 AA;

Query Match 55.0%; Score 38.5; DB 19; Length 2257;  
Best Local Similarity 54.5%; Pred. No. 6.7e+02;  
Matches 6; Conservative 2; Mismatches 2; Indels 1; Gaps 1;

QY 1 CXWLE-WLPC 10  
DB 1420 cqwevklwldc 1430

RESULT 10  
W05212  
ID W05212 standard; Protein; 2260 AA.

XX W05212;

XX 06-FEB-1997 (first entry)

XX Wheat acetyl-CoA carboxylase.

XX Acetyl-CoA carboxylase; oilseed; vegetable oil;  
KW arylloxyphenoxypionate; cyclohexanedione; herbicide resistance;  
KW monocot; transgenic plant; crop improvement; wheat.

XX Triticum aestivum var. Tam107, Hard Red Winter.

XX Key Location/Qualifiers  
FH MISC-difference 1980  
FT /note= "unidentified amino acid"

XX XX  
 PN W09632484-A2.  
 XX  
 PD 17-OCT-1996.  
 XX  
 PF 12-APR-1996; 96MO-US05095.  
 XX  
 PR 05-MAR-1996; 960S-0611546.  
 PR 14-APR-1995; 950S-0422560.  
 PR 06-JUN-1995; 95US-0468793.  
 XX  
 PA (ARCH-) ARCH DEV CORP.  
 PI Gornick P, Haselkorn R;  
 XX  
 DR WPI: 1996-485466/48.  
 DR N-PSDB; T43075.  
 XX  
 PT DNA encoding plant or cyanobacterial acetyl-CoA carboxylase enzyme  
 PT - useful for regulating oil content and for conferring and altering  
 PT resistance to herbicides which target acetyl-CoA carboxylase in  
 PT monocots  
 PS Claim 14; Page 202-220; 264pp; English.  
 CC  
 CC Wheat acetyl-CoA carboxylase (ACCase) (W05212) catalyzes the first  
 CC committed step in de novo fatty acid biosynthesis, the addition of  
 CC CO2 to acetyl-CoA to yield malonyl-CoA. Its amino acid sequence  
 CC was deduced from a complete gene sequence (T43075) isolated from a  
 CC wheat genomic library. Plant and cyanobacterial ACCases (W05205-  
 CC 12) can be used to modulate the oil content and herbicide resistance  
 CC of transgenic plants (esp. monocots). They may also be expressed in  
 CC transformed host (esp. cyanobacterial, E. coli or monocot plant)  
 CC cells and used to raise antibodies useful in the immunodetection of  
 CC ACCase.  
 CC  
 SX Sequence 2260 AA;  
 QY 1 CXWELR-WLPC 10  
 DB 1424 cqwvkrk1dc 1434  
 QY 1 CXWELR-WLPC 10  
 DB 1424 cqwvkrk1dc 1434  
 RESULT 11  
 Y60168 standard; Protein: 187 AA.  
 Y60168;  
 31-JAN-2000 (first entry)  
 Human endometrium tumour EST encoded protein 228.  
 Endometrium; human; tumour; cancer; anticancer; cytostatic; EST;  
 treatment; uterine; gene therapy; expressed sequence tag.  
 Homo sapiens.  
 DE19817948-A1.  
 21-OCT-1999.  
 17-APR-1998; 98DE-1017948.  
 17-APR-1998; 98DE-1017948.  
 (META-) METAGEN GES GENOMFORSCHUNG MBH.

PI Rosenthal A, Specht T, Hinzmann B, Schmitt A, Plarsky C, Dahl E;  
 XX WPI: 1999-591957/51.  
 DR N-PSDB; Z42055.  
 XX  
 PT New nucleic acid sequences expressed in uterine cancer tissues, and  
 PT derived polypeptides, for treatment of uterine and endometrial cancer  
 PT and identification of therapeutic agents  
 PS Claim 23; Page 366; 444pp; German.  
 CC  
 CC This invention describes novel human nucleic acid (cDNA) sequences (A),  
 CC that are highly expressed in uterine tumour tissue and which have  
 CC anticancer and cytostatic activity. (A) are used (i) for recombinant  
 CC expression of polypeptides (B) and (ii) to isolate complete genes. (B)  
 CC are used (i) to identify agents suitable for treatment of uterine or  
 CC endometrial cancer; (ii) directly for treating these forms of cancer  
 CC (including expression from gene therapy vectors) and (iii) for generation  
 CC of specific antibodies. (A) are identified by assembling ESTs (expressed  
 CC sequence tags) from a particular tissue type before comparison of  
 CC expression patterns. This allows a significantly longer fragment of the  
 CC gene to be revealed, so should reduce the number of failures associated  
 CC with the fact that ESTs from different libraries may represent different  
 CC parts of the same unknown gene, distorting the estimated frequency of  
 CC occurrence in a particular tissue. Y59941-Y60328 represent protein  
 CC fragments encoded by the human endometrium tumour cDNA library derived  
 CC EST fragments represented in Z41981-Z42121.  
 CC  
 SX Sequence 187 AA;  
 QY 1 CXWELR-WLPC 10  
 DB 62 cfwkawlhc 71  
 QY 1 CXWELR-WLPC 10  
 DB 62 cfwkawlhc 71  
 RESULT 12  
 W25945 standard; Protein: 421 AA.  
 W25945;  
 12-DEC-1997 (first entry)  
 Ob binding protein.  
 Ob binding protein.  
 Ob binding protein; obesity protein; Ob protein; appetite;  
 agonist; antagonist.  
 Synthetic.  
 Key Location/Qualifiers  
 MISC-difference 186 /note= "unidentified amino acid"  
 MISC-difference 195 /note= "unidentified amino acid"  
 MISC-difference 228 /note= "unidentified amino acid"  
 MISC-difference 391 /note= "unidentified amino acid"  
 W09720933-A2.  
 12-JUN-1997.  
 03-DEC-1996; 96MO-US18561.  
 06-DEC-1995; 95US-0568077.

PA (SCHE ) SCHERING CORP.  
 XX Altman SW, Bazan JF, Kastelein RA, Rock F;  
 PI WPI; 1997-332482/30.  
 DR  
 XX  
 PT Mammalian obesity protein variants - having alterations in helix  
 PT or/loop structures; used to treat obesity by decreasing appetite  
 PT or increasing metabolic rate in mammals  
 PS  
 XX Example; Page 35-36; 41pp; English.  
 XX  
 CC This polypeptide is an example of an Ob binding protein that can  
 CC be used in binding protein analyses of novel mammalian obesity (Ob)  
 CC protein mutants (see W25946-48). It can also be used in kits and  
 CC assay methods to screen compounds for interactions with Ob protein  
 CC mutants. The mutants are used as agonists or antagonists to treat  
 CC obesity by decreasing appetite or increasing metabolic rate in  
 CC mammals, especially humans.  
 SQ Sequence 421 AA;

Query Match 54.3%; Score 38; DB 18; Length 421;  
 Best Local Similarity 60.0%; Pred. No. 1.5e+02;  
 Matches 6; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
 QY 1 CXMELEWLP 10  
 I I I I I I  
 Db 344 csweaeglh 353

RESULT 13  
 W55884  
 ID W55884 standard; Protein; 551 AA.  
 AC W55884;  
 XX  
 AC 21-JUL-1998 (first entry)  
 DE Human CD33-like protein.  
 XX  
 DE Human CD33-like protein.  
 XX  
 KW Human; CD33; CD33-like protein; tumour; inflammatory disease;  
 KW Leukaemia; bone marrow; monocyte; haematopoietic; antibody.  
 XX  
 OS Homo sapiens.  
 XX  
 Key Location/Qualifiers  
 FT Peptide 1..15  
 FT /label= signal  
 FT Domain 16..422  
 FT /label= extracellular\_domain  
 FT Domain 423..464  
 FT /label= transmembrane\_domain  
 FT Domain 465..551  
 FT /label= intracellular\_domain  
 XX  
 PN W09806733-A1.  
 PD 19-FEB-1998.  
 XX  
 PF 09-AUG-1996; 96WO-US13007.  
 XX  
 PR 09-AUG-1996; 96WO-US13007.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 PI Gentz RL, Ni J, Rosen CA;  
 XX WPI; 1998-159451/14.  
 DR N-PSDB; V25979.  
 XX  
 PT New nucleic acid nearly identical to sequence encoding CD33-like

PT protein - useful in, e.g. diagnosis of tumour or inflammatory  
 PT disease and purging bone marrow monocytic haematopoietic cells from  
 PT leukaemia patients  
 XX  
 PS Claim 10; Fig 1; 83pp; English.  
 XX  
 CC The present sequence represents a CD33-like protein. The cDNA encoding  
 CC the present sequence was obtained by sequencing the HMOCD14 cDNA clone  
 CC contained in ATCC Deposit No. 97521. An isolated antibody that binds  
 CC specifically to the CD33-like protein may be used for the detection of  
 CC the CD33-like protein or its mRNA, and so is useful for, e.g. diagnosing  
 CC a tumour or inflammatory disease. The antibody (especially an  
 CC immunotoxin), can also be used to remove or deplete haematopoietic cells  
 CC expressing the CD33-like protein antigen, which can be used to purge  
 CC bone marrow monocytic haematopoietic cells obtained from a leukaemia  
 CC patient, which can subsequently be reinfused into a patient previously  
 CC subjected to myelablative chemotherapy. The antibody can also be used  
 CC as an antagonist to inhibit the CD33-like protein receptor signalling  
 CC pathway, useful for inhibiting the growth or selective killing of tumour  
 CC cells.  
 SQ Sequence 551 AA;

Query Match 54.3%; Score 38; DB 19; Length 551;  
 Best Local Similarity 60.0%; Pred. No. 2e+02;  
 Matches 6; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
 QY 1 CXMELEWLP 10  
 I I I I I I  
 Db 342 csweaeglh 351

RESULT 14  
 R93121  
 ID R93121 standard; Protein; 606 AA.  
 AC R93121;  
 XX  
 AC 10-OCT-1996 (first entry)  
 DE Tllapia prolactin receptor (mature form).  
 XX  
 DE Tllapia prolactin receptor (mature form).  
 XX  
 KW Fish prolactin; tllprl; receptor; hormone; agonist; antagonist;  
 KW reproductive cycle synchronisation; teleost; bony fish;  
 KW Cypriniformes.  
 XX  
 OS Oreochromis niloticus.  
 XX  
 Key Location/Qualifiers  
 FT Protein 1..606  
 FT /label= prolactin\_receptor  
 FT Domain 1..210  
 FT /label= extracellular  
 FT /note= "comprises 5 Cys residues"  
 FT Binding-site 192..196  
 FT /label= ligand binding motif  
 FT /note= "conforms to Trp-Ser-Xaa-Trp-Ser consensus"  
 FT Modified-site 68..70  
 FT /label= glycosylation\_site  
 FT /note= "potential"  
 FT Modified-site 77..79  
 FT /label= glycosylation\_site  
 FT /note= "potential"  
 FT Domain 211..234  
 FT /label= transmembrane  
 FT Domain 235..606  
 FT /label= cytoplasmic  
 FT /note= "slightly longer than mammalian PRL receptor  
 FT cytoplasmic domain"  
 FT Region 245..250  
 FT /label= Box\_1  
 FT /note= "proline-rich region highly conserved among

receptors for cytokines, growth hormones and prolactin and is critical for signal transduction"

FR2724181-A1.

08-MAR-1996.

01-SEP-1994; 94FR-0010535.

01-SEP-1994; 94FR-0010535.

(INRM ) INSERM INST NAT SANTE & RECH MEDICALE.

Ederly M, Kelly PA, Prunet P, Sandra O;

WPI: 1996-153124/16.

N-PSDB; T17141.

Fish prolactin receptor - useful for detecting prolactin agonists and antagonists.

Claim 5; Page 20-21; 35pp; French.

A cDNA sequence coding for a fish prolactin (PRL) receptor was isolated from a Oreochromis niloticus (tilapia) kidney cDNA expression library following screening with radioactive tPRL. The cDNA insert was found to contain an open reading frame for a 630 amino acid protein. The mature protein (606 amino acids) has estimated mol. wt. 68.2 kDa and isoelectric point 5.53. Transformed eukaryotic cells expressing the PRL receptor are useful for identifying agonists and antagonists of PRL which have potential applications in fish farming, e.g. for synchronising reproductive cycles. The present sequence is that of the mature PRL receptor from tilapia.

Sequence 606 AA;

Query Match 54.3%; Score 38; DB 17; Length 606;  
Best Local Similarity 66.7%; Pred. No. 2.2e+02;  
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

3 WELEWLPKA 11  
|||:||||

368 wekeampca 376

RESULT 15  
R93120  
R93120 standard; Protein; 630 AA.

R93120;

09-OCT-1996 (first entry)

Tilapia prolactin receptor precursor.

Fish prolactin; tPRL; receptor; hormone; agonist; antagonist;  
KW reproductive cycle synchronisation; teleost; bony fish;  
Cypriiniformes.

Oreochromis niloticus.

Key Location/Qualifiers  
Protein 25..630  
FT /label= prolactin\_receptor  
FT 25..234  
FT /label= extracellular  
FT /note= "comprises 5 Cys residues"  
FT Binding-site 216..220  
FT /label= ligand binding motif  
FT /note= "conforms to Trp-Ser-Xaa-Trp-Ser consensus"

Modified-site 92..94  
FT /label= glycosylation\_site  
FT /note= "potential"  
FT Modified-site 101..103  
FT /label= glycosylation\_site  
FT /note= "potential"  
FT Domain 235..258  
FT /label= transmembrane  
FT Domain 259..630  
FT /label= cytoplasmic  
FT /note= "slightly longer than mammalian PRL receptor cytoplasmic domain"

Region 269..274  
FT /label= Box\_1  
FT /note= "proline-rich region highly conserved among  
FT receptors for cytokines, growth hormones and  
FT prolactin and is critical for signal  
FT transduction"

FR2724181-A1.

08-MAR-1996.

01-SEP-1994; 94FR-0010535.

01-SEP-1994; 94FR-0010535.

(INRM ) INSERM INST NAT SANTE & RECH MEDICALE.

Ederly M, Kelly PA, Prunet P, Sandra O;

WPI: 1996-153124/16.

N-PSDB; T17141.

Fish prolactin receptor - useful for detecting prolactin agonists and antagonists.

Example; Page 15-16; 35pp; French.

A cDNA sequence coding for a fish prolactin (PRL) receptor was isolated from a Oreochromis niloticus (tilapia) kidney cDNA expression library following screening with radioactive tPRL. The cDNA insert was found to contain an open reading frame for a 630 amino acid protein. The mature protein (606 amino acids) has estimated mol. wt. 68.2 kDa and isoelectric point 5.53. Transformed eukaryotic cells expressing the PRL receptor are useful for identifying agonists and antagonists of PRL which have potential applications in fish farming, e.g. for synchronising reproductive cycles. The present sequence is that of the precursor for the PRL receptor from tilapia.

Sequence 630 AA;

Query Match 54.3%; Score 38; DB 17; Length 630;  
Best Local Similarity 66.7%; Pred. No. 2.2e+02;  
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

3 WELEWLPKA 11  
|||:||||

392 wekeampca 400

Search completed: April 11, 2001, 12:58:00  
Job time: 138 sec



Fri Apr 20 13:16:29 2001

us-09-722-440-8.rag

---



3  
.  
.  
.  
.